

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 04:02:20 ; Search time 125 Seconds
(without alignments)

431.733 Million cell updates/sec

Title: US-09-713-687A-1

Perfect score: 1052

Sequence: 1 MSINPKPQLTCNTNRPPQ.....CSFSIFLLALLSLCTVPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:*

1: geneseqp1980s.*

2: geneseqp1980s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	998	94.9	191	2	AAR44010 Hepatitis
2	998	94.9	191	2	AAR92938 Hepatitis
3	998	94.9	191	2	AAR92941 Hepatitis
4	998	94.9	191	2	AAR92937 Hepatitis
5	998	94.9	191	2	AAR92939 Hepatitis
6	998	94.9	191	2	AAR92940 Hepatitis
7	998	94.9	191	3	AAY94411 Human hep
8	998	94.9	967	2	AAR79222 PHCV141-e
9	998	94.9	1006	2	AAR12715 HCV Genom
10	998	94.9	1648	2	AAR79221 PHCV176-e
11	998	94.9	2984	4	AAR00449 Hepatitis
12	998	94.9	2984	4	AAR00447 Hepatitis
13	998	94.9	2984	4	AAR00442 Hepatitis
14	998	94.9	3011	2	AAR40119 HCV Genom
15	998	94.9	3011	2	AAR40120 HCV Genom
16	998	94.9	3011	2	AAR79232 HCV sequ
17	998	94.9	3011	2	AAR77397 Hepatitis
18	998	94.9	3011	2	AAR77398 Hepatitis
19	998	94.9	3011	2	AAR98020 Infectiou
20	998	94.9	3011	4	AAB59173 Protein e
21	998	94.9	3011	4	AAB31169 Amino aci
22	998	94.9	3011	5	AAR94597 HCV polyp
23	998	94.9	3011	5	AAR79221 Hepatitis
24	998	94.9	3011	5	AAR19888 Hepatitis
25	998	94.9	3011	6	ABP71460 Amino aci

26	998	94.9	3011	7	ABW00339	Hepatitis
27	998	94.9	3012	5	AAU99289	Hepatitis
28	998	94.9	3012	6	ABU61848	HCV H77 c
29	998	94.9	3180	6	ABG73195	MKO-Z vir
30	998	94.9	3208	7	ADD67948	Hepatitis
31	994	94.5	3011	5	AAU99290	Hepatitis
32	994	94.5	3011	6	AAO26784	Protein d
33	994	94.5	3011	6	ABU61849	HCV-H. 8/
34	993	94.4	3835	4	ABU61167	HCV-H. 8/
35	992	94.3	191	2	AAR92936	Hepatitis
36	992	94.3	3011	2	AAR22154	Hepatitis
37	991	94.2	502	2	AAR67589	Hepatitis
38	990	94.1	191	3	AAU82993	Hepatitis
39	990	94.1	2816	2	AAAR34009	HCV-1 pol
40	990	94.1	2955	2	AAU14975	Amino aci
41	990	94.1	2955	3	AAU18541	Polyprote
42	990	94.1	3011	2	AAR21519	Compiled
43	990	94.1	3011	2	AAU90931	Hepatitis
44	990	94.1	3011	2	AAW34480	HCV polyp
45	990	94.1	3011	2	AAW40038	HCV polyp

ALIGNMENTS

RESULT 1
AAR44010

ID AAR44010 standard; protein; 191 AA.

XX AC AAR44010;

XX DT 25-MAR-2003 (revised)

DT 12-MAY-1994 (first entry)

XX DE Hepatitis C Virus core protein.

XX KW HCV; non-A, non-B hepatitis virus; NANBHV; immunodominant region;

KW capsid protein; vaccine; antigen.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

FT Peptide 1..45

FT /label= S42G

FT /note= "peptide 1"

FT Region 1..45

FT /label= immunodominant region

FT /note= "novel antigenic peptides contain sequences from this region"

FT Peptide 38..81

FT /label= P42Y

FT /note= "peptide 2"

FT Peptide 77..115

FT /label= R40R

FT /note= "peptide 3"

XX PN EP569309-A1.

XX 10-NOV-1993.

XX PP 06-MAY-1993; 93EP-00420183.

XX PR 06-MAY-1992; 92PR-00005763.

XX (INMR) BIO MERIEUX.

XX PI Dalbon P, Jolivet M;

XX DR WPI; 1993-353383/45.

XX PT New antigenic polypeptide(s) from hepatitis C virus - derived from N-

PT terminal region of core protein, and related antibodies, useful in

PT diagnosis, vaccination and treatment.


```

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
DB 181 LLSCLTVPASA 191

RESULT 6
ID AAR92940
AC AAR92940
XX AAR92940;
DT 02-OCT-1996 (first entry)
DE Hepatitis C virus isolate S18 core protein.
KW HCV; EL: envelope 1; core protein; HCV genotyping; antibody; vaccine;
  hepatitis.
OS Hepatitis C virus.
PN WO9605315-A2.
XX WO9605315-A2.
PD 22-FEB-1996.
XX 15-AUG-1995; 95WO-US010398.
XX 15-AUG-1994; 94US-00290665.
XX (USSH ) US SEC DEPT HEALTH.
PA Bukh J, Miller RH, Purcell RH;
PI WPI; 1996-139709/14.
DR N-PSDB; AAT11614.
XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
  determine HCV genotype and as vaccines against HCV infection.
PS Claim 4; Page 184; 340pp; English.
XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
  isolates. Isolated cDNA sequences are used for the prodn. of primers
  useful for detecting the presence of HCV in a sample, the primers are
  also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
  in vaccines for immunising against HCV infection. The proteins may also
  be used to detect antibodies against HCV in serum, saliva, lymphocytes or
  other mononuclear cells. The antibodies may be used in the prevention of
  HCV infection
XX Sequence 191 AA;
  Query Match 94.9%; Score 998; DB 2; Length 191;
  Best Local Similarity 96.3%; Pred. No. 1.5e-87;
  Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSERSNPRG 60
DB 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSERSQPRG 60
QY 61 RNPPIKARRPDGRTWANGPYWPPLYNGEGCGAGWLLSPRGRSPSWGPTDPRRSRNIG 120
DB 61 RQPIPKARRPGRGTWANGPYWPPLYNGEGCGAGWLLSPRGRSPSWGPTDPRRSRNIG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191

```

```

DB 181 LLSCLTVPASA 191
  |||||
RESULT 7
ID AAY94411
XX AAY94411 standard; peptide; 191 AA.
AC AAY94411;
XX 11-SEP-2000 (first entry)
DE Human hepatitis C virus core protein.
KW Human; hepatitis C virus; HCV; immunogen; antigen; antibody; virucide;
  hepatotropic; anti-inflammatory; virus detection; vaccine.
OS Hepatitis C virus.
XX WO200031130-A1.
PD 02-JUN-2000.
XX 19-NOV-1999; 99WO-IB001933.
XX 20-NOV-1998; 98US-00196155.
XX (INMR ) BIO MERIEUX.
PA Dalbon P, Jolivet M, Jolivet-Reynaud C;
PI WPI; 2000-411934/35.
XX Polypeptides that bind to anti-hepatitis C virus antibodies, useful for
  diagnosing and preventing hepatitis C infections.
PS Disclosure; Fig 1; 50pp; English.
XX The present sequence is the core protein of the human hepatitis C virus
  (HCV). Three long polypeptide fragments of about 40 amino acids in the
  120 N-terminal amino acids of this sequence were synthesised and their
  reactivity towards HCV-positive sera was evaluated in an ELISA test. The
  reactivity of the first peptide, S42G, which extends from serine 2 up to
  glycine 45, was substantially greater than the reactivity of the other
  peptides. Peptide S42G represents an immunodominant region containing
  conformational type epitopes and linear type epitopes. S42G manifests an
  immunoreactivity with all the sera of individuals or blood samples
  infected with HCV and which possess antibodies directed against the core
  protein. An amino acid of the S42G sequence may be substituted for
  homologous amino acids and side chains and peptide bonds may also be
  modified. For example, L-amino acids may be replaced by D-amino acids,
  amine groups may be acetylated, and so on. The native antigenic sequence
  and its antigenic derivatives may be used for detection of human
  hepatitis C viruses and for raising antibodies against the virus
XX Sequence 191 AA;
  Query Match 94.9%; Score 998; DB 3; Length 191;
  Best Local Similarity 96.3%; Pred. No. 1.5e-87;
  Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSERSNPRG 60
DB 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSERSQPRG 60
QY 61 RNPPIKARRPDGRTWANGPYWPPLYNGEGCGAGWLLSPRGRSPSWGPTDPRRSRNIG 120
DB 61 RQPIPKARRPGRGTWANGPYWPPLYNGEGCGAGWLLSPRGRSPSWGPTDPRRSRNIG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180

```

QY 181 LLCLTVPASA 191
 Db 181 LLCLTVPASA 191

RESULT 8

AAR79222
 ID AAR79222 standard; protein; 967 AA.

XX AC AAR79222;

XX 08-DEC-1995 (first entry)

XX pHCV141-encoded sequence.

XX pHCV141; amyloid precursor protein; APP; hepatitis C virus; HCV; E1; E2;
 KW fusion protein; HEK-293; vaccine; vector; pRC/CMV; protein secretion;
 glycosylation.

XX OS Synthetic.

XX WO9520664-A2.

XX 03-AUG-1995.

XX 27-JAN-1995; 95WO-US001087.

XX 28-JAN-1994; 94US-00188281.

XX (ABBO) ABBOTT LAB.

XX Watanabe S, Yamaguchi J, Desai SM, Devare SG;

XX WPI; 1995-275449/36.

XX New mammalian expression systems for HCV proteins - express fusion
 PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
 PT protein.

XX Disclosure; Page 59-62; 89pp; English.

XX RNA from serum or plasma of a chimpanzee infected with HCV was converted
 CC to cDNA and PCR amplified using primers based on HCV sequences.⁷
 CC Adjacent HCV DNA fragments were generated which together encoded the HCV
 CC sequence given in AAR79222. Fragments from 2 clones, pHCV141 and pHCV150
 CC (see AAR79223), were combined to generate pHCV176 (see AAR79221). This
 CC was used to construct APP-HCV-E1+E2 fusion proteins that were
 CC glycosylated and secreted from HEK-293 transfectants

XX Sequence 967 AA;

Query Match 94.9%; Score 998; DB 2; Length 967;
 Best Local Similarity 96.3%; Pred. No. 9.7e-87;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCNTNRRPQDVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSRSNPRG 60
 Db 1 MSTNPKPQRTKXNTNRRPQDVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSRSQPRG 60
 QY 61 RRNPPIKARRPDGRTWANPGYPWPPLYNCGCGWAGWLLSPRGSRPSWGPTDPMWRSRNLG 120
 Db 61 RRQPIKARRPEGRTWAPQGYWPPLYNCGCGWAGWLLSPRGSRPSWGPTDPMWRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPLVGAIPGGAAALAHGVRVLEDGVNATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAIPGGAAALAHGVRVLEDGVNATGNLPGCSFSIFLLA 180
 QY 181 LLCLTVPASA 191
 Db 181 LLCLTVPASA 191

RESULT 9

AAR79221

ID AAR79221 standard; protein; 1648 AA.

AAW12715

ID AAW12715 standard; protein; 1006 AA.

XX AC AAW12715;

XX 17-OCT-2003 (revised)

XX 09-MAY-1997 (first entry)

XX HCV genome type 1a(H) amino acid residues 1-1006.

XX HCV; influenza virus; vaccine; fusion protein; immunogen; core protein;
 KW NS1; plasmid pRIT14129.

XX Hepatitis C virus; genotype 1a(H).

XX Key Location/Qualifiers

FT Region 2..166

FT /label= NS1

FT /note= "amino acids 2-166 are utilised in a novel NS1-
 Core fusion protein"

XX WO9701640-A2.

XX 16-JAN-1997.

XX 20-JUN-1996; 96WO-EP002764.

XX 29-JUN-1995; 95GB-00013261.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cabazon Silva T, Momin FM, Garoon NMC;

XX WPI; 1997-100211/09.

XX New vaccines against hepatitis C virus - comprising QS21, 3D-MPL, an oil
 PT in water emulsion and a HCV core or envelope protein or immunogenic
 PT deriv.

XX Example 1; Page 15-16; 20pp; English.

XX Residues 1-1006 of the hepatitis C virus (HCV) genome type 1a(H) are
 CC shown in AAW12715. Residues 2-166 (from the core protein) have been
 CC utilised in a novel fusion protein (see also AAW12714) with influenza
 CC virus NS1. The fusion protein can be formulated into novel vaccines
 CC against HCV. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 1006 AA;

Query Match 94.9%; Score 998; DB 2; Length 1006;
 Best Local Similarity 96.3%; Pred. No. 1e-86;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCNTNRRPQDVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSRSNPRG 60
 Db 1 MSTNPKPQRTKXNTNRRPQDVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSRSQPRG 60
 QY 61 RRNPPIKARRPDGRTWANPGYPWPPLYNCGCGWAGWLLSPRGSRPSWGPTDPMWRSRNLG 120
 Db 61 RRQPIKARRPEGRTWAPQGYWPPLYNCGCGWAGWLLSPRGSRPSWGPTDPMWRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPLVGAIPGGAAALAHGVRVLEDGVNATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAIPGGAAALAHGVRVLEDGVNATGNLPGCSFSIFLLA 180
 QY 181 LLCLTVPASA 191
 Db 181 LLCLTVPASA 191

```

XX AAR79221;
AC
XX
XX 08-DEC-1995 (first entry)
DT
XX
XX pHCV176-encoded sequence.
DE
XX
XX pHCV176; amyloid precursor protein; APP; hepatitis C virus; HCV; E1; E2;
KW fusion protein; HEK-293; vaccine; vector; pRC/CMV; protein secretion;
KW glycosylation.
XX
XX Synthetic.
OS
XX
XX W09520664-A2.
PN
XX
XX 03-AUG-1995.
PD
XX
XX 27-JAN-1995; 95WO-US001087.
PF
XX
XX 28-JAN-1994; 94US-00188281.
PR
XX
XX (ABBO ) ABBOTT LAB.
PA
XX
XX Watanabe S, Yamaguchi J, Desai SM, Devare SG;
PI
XX
XX WPI; 1995-275449/36.
DR
XX
XX New mammalian expression systems for HCV proteins - express fusion
PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
PT protein.
XX
XX Disclosure; Page 53-59; 89pp; English.
PS
XX
XX RNA from serum or plasma of a chimpanzee infected with HCV was converted
CC to cDNA and PCR amplified using primers based on HCV sequences. 7
CC Adjacent HCV DNA fragments were generated which together encoded the HCV
CC sequence given in AAR79232. Fragments from 2 clones, pHCV141 (see
CC AAR79222) and pHCV150 (see AAR79223), were combined to generate pHCV176.
CC This was used to construct APP-HCV-E1+E2 fusion proteins that were
CC glycosylated and secreted from HEK-293 transfectants
XX
XX Sequence 1648 AA;
SQ
Query Match 94.9%; Score 998; DB 2; Length 1648;
Best Local Similarity 96.3%; Pred. No. 1.8e-86;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
Db 1 MSTNPKPQKTKRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RNPPIKARRPDGRTWANGYPWPYLYGNEGCGWAGWLLSPRGRSPSWGPTDPRRSRLG 120
Db 61 RQPIPKARRPDGRTWANGYPWPYLYGNEGCGWAGWLLSPRGRSPSWGPTDPRRSRLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVDGYNVYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVDGYNVYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191
RESULT 11
AAE00449
ID AAE00449 standard; protein; 2984 AA.
XX
XX AAE00449;
AC
XX
XX 19-JUN-2001 (first entry)
DT
XX
XX Hepatitis C virus H77C protein lacking HVR1 from chimpanzee #96A008.
DE

```

```

XX
KW Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
KW antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutant;
KW HCV infection; viral replication; passive immunoprophylaxis; mutein.
XX
XX Hepatitis C virus.
OS
XX
XX Key Location/Qualifiers
FH 192-383
FT Region /label= E1_protein
FT Misc-difference 253
FT /note= "Wild type Lys substituted with Asn"
FT Region 384..746
FT /label= E2_protein
FT /note= "E2 protein lacks hypervariable region 1"
FT Misc-difference 487
FT /note= "Wild type Val substituted with Met"
FT Misc-difference 1116
FT /note= "Wild type Arg substituted with His"
FT Misc-difference 1283
FT /note= "Wild type Tyr substituted with His"
XX
XX W0200121807-A1.
PN
XX
XX 29-MAR-2001.
PD
XX
XX 22-SEP-2000; 2000WO-US025987.
PF
XX
XX 23-SEP-1999; 99US-0155823P.
PR
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Forns X, Bukh J, Emerson SU, Purcell RH;
PI
XX
XX WPI; 2001-266076/27.
DR
XX
XX N-PSDB; AAD03808.
XX
Novel nucleic acid molecules that encode hepatitis C virus envelope 2
PT protein lacking all or part of hypervariable region 1 of envelope
PT protein, useful as vaccine components for treating or preventing HCV
PT infections.
XX
XX Example 5; Page; 80pp; English.
XX
The present sequence is Hepatitis C virus (HCV) H77C protein from
CC chimpanzee #96A008 which lacks the hypervariable region one (HVR1) of HCV
CC envelope 2 (E2). The HCV E2 protein lacking HVR1 (H77C(HVR1)) DNA is
CC useful for producing infectious HCV and chimeric HCV viruses which are
CC are useful for identifying cell lines capable of supporting the
CC replication of viruses. The infectious HCV and HVR1- chimeric HCV are
CC used in the production of attenuated or inactivated vaccines which are
CC useful for treating or preventing HCV in a mammal by immunisation. The
CC host calls expressing the H77C(HVR1)DNA is useful as an immunogen to
CC stimulate a protective immune response to HCV. The immunogens are useful
CC for producing protective antibodies to HCV. The antibodies produced are
CC used in passive immunoprophylaxis for treatment of diseases caused by HCV
CC in animals, especially humans. The H77C(HVR1) DNA is also useful in gene
CC therapy. Note: The present sequence is not found in the specification but
CC is derived from Hepatitis C virus envelope protein 2 lacking
CC hypervariable region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in
CC figure 1
XX
XX Sequence 2984 AA;
SQ
Query Match 94.9%; Score 998; DB 4; Length 2984;
Best Local Similarity 96.3%; Pred. No. 3.6e-86;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
Db 1 MSTNPKPQKTKRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
QY 61 RNPPIKARRPDGRTWANGYPWPYLYGNEGCGWAGWLLSPRGRSPSWGPTDPRRSRLG 120

```


PT Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 PT protein lacking all or part of hypervariable region 1 of envelope
 PT protein, useful as vaccine components for treating or preventing HCV
 PT infections.
 XX Claim 1; Fig 1G-H; 80pp; English.
 XX
 XX The present sequence is infectious hepatitis C virus (HCV) of genotype 1a
 CC clone [H77C(HVR1)] which lacks the hypervariable region one (HVR1) of HCV
 CC envelope 2 (E2). The HCV E2 protein lacking HVR1 DNA is useful for
 CC producing infectious HCV and chimeric HCV viruses which are useful
 CC for identifying cell lines capable of supporting the replication of
 CC viruses. The infectious HCV and HVR1-chimeric HCV are used in the
 CC production of attenuated or inactivated vaccines which are useful for
 CC treating or preventing HCV in a mammal by immunisation. The host cells
 CC expressing the H77C(HVR1) DNA is useful as an immunogen to stimulate a
 CC protective immune response to HCV. The immunogens are useful for
 CC producing protective antibodies to HCV. The antibodies produced are used
 CC in passive immunoprophylaxis for treatment of diseases caused by HCV in
 CC animals, especially humans. The H77C(HVR1) DNA is also useful in gene
 CC therapy
 XX
 SQ Sequence 2984 AA;

Query Match 94.9%; Score 998; DB 4; Length 2984;
 Best Local Similarity 96.3%; Pred. No. 3.6e-86;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MSTNPKPQRLTCRNTNRRPQDVKPPGGQIVGGVYLLPRRGLGVRATRTKTSERSNPRG 60
 DB 1 MSTNPKPQRTKRNTNRRPQDVKPPGGQIVGGVYLLPRRGLGVRATRTKTSERSQPRG 60
 QY 61 RRPNTPKARRPDGRTWANPGYWPPLYGNEGCGWAGLLSPRGSRSWGPTDPWRRSRNLG 120
 DB 61 RRQPIPKARRREGRTWAQPGYWPPLYGNEGCGWAGLLSPRGSRSWGPTDPWRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTVPASA 191
 DB 181 LLSCLTVPASA 191

RESULT 14
 AAR40119
 ID AAR40119 standard; protein; 3011 AA.
 XX
 AC AAR40119;
 XX
 DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 27-JAN-1994 (first entry)
 XX
 DE HCV genomic amino acid sequence isolated from infected chimpanzee CO.
 XX Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
 KW human growth hormone; HGH; secretion signal; fusion protein; vaccine.
 XX
 OS Hepatitis C virus; Virus.
 XX
 PN WO9315193-A1.
 XX
 PD 05-AUG-1993.
 XX
 PF 29-JAN-1993; 93WO-US000907.
 XX
 PR 31-JAN-1992; 92US-00830024.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Casey JM, Bode SL, Zeck BJ, Yamaguchi J, Frail DE, Desai SM;

PI Devare SG;
 XX
 DR WPI; 1993-258673/32.
 XX
 PT New plasmid pHCV-162 is a mammalian expression systems for HCV proteins
 PT - useful for diagnosing HCV infection and as vaccines for preventing HCV
 PT infection.
 XX
 PT Example 1; Page 29-39; 100pp; English.
 PS
 XX RNA was isolated from the serum of a chimpanzee (designated "CO")
 CC experimentally infected with HCV and cDNA was prepared from it. The cDNA
 CC was PCR amplified using specific primers with sequences based on the
 CC prototype HCV-1 cDNA sequence (GENBANK M62321). Further amplification
 CC using nested primers resulted in 7 adjacent HCV DNA fragments which could
 CC be assembled into a full-length sequence. The DNA sequence was determined
 CC and translated into the genomic amino acid sequence. Comparison of the CO
 CC genomic amino acid sequence with that from HCV-1 showed 98 amino acid
 CC differences. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24
 CC -OCT-2003 to standardise OS field)
 XX
 SQ Sequence 3011 AA;

Query Match 94.9%; Score 998; DB 2; Length 3011;
 Best Local Similarity 96.3%; Pred. No. 3.6e-86;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MSTNPKPQRLTCRNTNRRPQDVKPPGGQIVGGVYLLPRRGLGVRATRTKTSERSNPRG 60
 DB 1 MSTNPKPQRTKRNTNRRPQDVKPPGGQIVGGVYLLPRRGLGVRATRTKTSERSQPRG 60
 QY 61 RRPNTPKARRPDGRTWANPGYWPPLYGNEGCGWAGLLSPRGSRSWGPTDPWRRSRNLG 120
 DB 61 RRQPIPKARRREGRTWAQPGYWPPLYGNEGCGWAGLLSPRGSRSWGPTDPWRRSRNLG 120
 QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
 DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTVPASA 191
 DB 181 LLSCLTVPASA 191

RESULT 15
 AAR40120
 ID AAR40120 standard; protein; 3011 AA.
 XX
 AC AAR40120;
 XX
 DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 27-JAN-1994 (first entry)
 XX
 DE HCV genomic amino acid sequence isolated from infected human LG.
 XX Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
 KW human growth hormone; HGH; secretion signal; fusion protein; vaccine.
 XX
 OS Hepatitis C virus; Virus.
 XX
 PN WO9315193-A1.
 XX
 PD 05-AUG-1993.
 XX
 PF 29-JAN-1993; 93WO-US000907.
 XX
 PR 31-JAN-1992; 92US-00830024.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Casey JM, Bode SL, Zeck BJ, Yamaguchi J, Frail DE, Desai SM;
 PI Devare SG;


```
XX WPI; 1993-258673/32.
DR
XX
XX New plasmid pHCV-162 is a mammalian expression systems for HCV proteins
PT - useful for diagnosing HCV infection and as vaccines for preventing HCV
PT infection.
XX
XX
XX Example 1; Page 39-49; 100pp; English.
PS
XX
XX RNA was isolated from the plasma of a HCV seropositive human (designated
CC "LG") and cDNA was prepared from it. The cDNA was PCR amplified using
CC specific primers with sequences based on the prototype HCV-1 cDNA
CC sequence (GENBANK M62321). Further amplification using nested primers
CC resulted in 7 adjacent HCV DNA fragments which could be assembled into a
CC full-length sequence. The DNA sequence was determined and translated into
CC the genomic amino acid sequence. Comparison of the LG genomic amino acid
CC sequence with that from HCV-1 showed 134 amino acid differences. (Updated
CC on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to
CC standardise OS field)
XX
XX Sequence 3011 AA;
SQ
Query Match 94.9%; Score 998; DB 2; Length 3011;
Best Local Similarity 96.3%; Pred. No. 3.6e-86;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSNPRG 60
Db |||||
1 MSTNPKPQRTKNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
QY 61 RRNPPIKARRPDGRTWANPGYWPFLYNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
Db |||||
61 RRQPIKARRPEGTWNAQPGYWPFLYNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDCGVNYATGNLPGCSFSIFLLA 180
Db |||||
121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDCGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLITVPASA 191
Db |||||
181 LLSCLITVPASA 191
```

Search completed: August 25, 2004, 04:25:12
Job time : 130 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 04:23:08 ; Search time 32 Seconds
(without alignments)
308.143 Million cell updates/sec

Title: US-09-713-687A-1

Perfect score: 1052

Sequence: 1 MSTNPKPQRLTCNTRRPQ.....CSFSIFLLALLSCLIVPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/6CTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	998	94.9	191	2	US-08-290-665A-156
2	998	94.9	191	2	US-08-290-665A-157
3	998	94.9	191	2	US-08-290-665A-158
4	998	94.9	191	2	US-08-290-665A-159
5	998	94.9	191	2	US-08-290-665A-160
6	998	94.9	191	3	US-08-380-160-3
7	998	94.9	191	5	PCT-US95-10398-156
8	998	94.9	191	5	PCT-US95-10398-157
9	998	94.9	191	5	PCT-US95-10398-158
10	998	94.9	191	5	PCT-US95-10398-159
11	998	94.9	191	5	PCT-US95-10398-160
12	998	94.9	450	4	US-08-635-886C-179
13	998	94.9	450	4	US-08-635-886C-180
14	998	94.9	450	4	US-08-974-690C-179
15	998	94.9	450	4	US-08-974-690C-180
16	998	94.9	967	1	US-08-188-281B-13
17	998	94.9	967	1	PCT-US94-07280-13
18	998	94.9	967	5	PCT-US95-01087-13
19	998	94.9	1648	1	US-08-188-281B-12
20	998	94.9	1648	5	PCT-US94-07280-12
21	998	94.9	1648	5	PCT-US95-01087-12
22	998	94.9	3011	1	US-08-188-281B-1
23	998	94.9	3011	1	US-08-433-552-1
24	998	94.9	3011	1	US-08-433-552-2
25	998	94.9	3011	2	US-08-710-637-1
26	998	94.9	3011	2	US-08-710-637-2
27	998	94.9	3011	3	US-09-014-416-1

28 998 94.9 3011 4 US-10-104-966-1 Sequence 1, Appli
29 998 94.9 3011 4 US-09-952-572-9 Sequence 9, Appli
30 998 94.9 3011 5 PCT-US93-00907-1 Sequence 1, Appli
31 998 94.9 3011 5 PCT-US93-00907-2 Sequence 2, Appli
32 998 94.9 3011 5 PCT-US94-07280-1 Sequence 1, Appli
33 998 94.9 3011 5 PCT-US95-01087-1 Sequence 1, Appli
34 998 94.9 3012 3 US-08-811-566-2 Sequence 2, Appli
35 998 94.9 3012 4 US-09-034-756-2 Sequence 2, Appli
36 998 94.5 3011 3 US-08-811-566-20 Sequence 20, Appli
37 994 94.3 3011 4 US-09-034-756-20 Sequence 20, Appli
38 992 94.3 191 2 US-08-290-665A-155 Sequence 155, App
39 992 94.3 191 5 PCT-US95-10398-155 Sequence 155, App
40 990 94.1 191 4 US-09-763-260-4 Sequence 36, Appli
41 990 94.1 3011 1 US-08-440-103-36 Sequence 36, Appli
42 990 94.1 3011 1 US-08-440-542-36 Sequence 36, Appli
43 990 94.1 3011 1 US-07-910-760-10 Sequence 10, Appli
44 990 94.1 3011 1 US-08-440-519-10 Sequence 10, Appli
45 990 94.1 3011 1 US-08-231-368-36 Sequence 36, Appli

ALIGNMENTS

RESULT 1
US-08-290-665A-156
; Sequence 156, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 156:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: hom sapiens
; INDIVIDUAL ISOLATE: US11
; US-08-290-665A-156

Query Match 94.9%; Score 998; DB 2; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;

Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
 |||||
 Db 1 MSTNPKPQKTKRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
 |||||

Qy 61 RNNIPKARRDPGRWTWANGYPWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
 |||||
 Db 61 RQQIPKARRPEGRWTAQPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
 |||||

Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
 |||||
 Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
 |||||

Qy 181 LLSCLTVPASA 191
 |||||
 Db 181 LLSCLTVPASA 191
 |||||

RESULT 2

US-08-290-665A-157
 ; Sequence 157, Application US/08290665A
 ; Patent No. 5892852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R. H. AND
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 157:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 191 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; ORIGINAL SOURCE:
 ; ORGANISM: hom sapiens
 ; INDIVIDUAL ISOLATE: S14
 ; US-08-290-665A-157

Query Match 94.9%; Score 998; DB 2; Length 191;
 Best Local Similarity 96.3%; Pred. No. 5.8e-92;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
 |||||

Db 1 MSTNPKPQKTKRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
 |||||

Qy 61 RNNIPKARRDPGRWTWANGYPWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
 |||||

Db 61 RQQIPKARRPEGRWTAQPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
 |||||

Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
 |||||

Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
 |||||

Qy 181 LLSCLTVPASA 191
 |||||

Db 181 LLSCLTVPASA 191
 |||||

RESULT 3

US-08-290-665A-158
 ; Sequence 158, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R. H. AND
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 158:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 191 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; ORIGINAL SOURCE:
 ; ORGANISM: hom sapiens
 ; INDIVIDUAL ISOLATE: SW1
 ; US-08-290-665A-158

Query Match 94.9%; Score 998; DB 2; Length 191;
 Best Local Similarity 96.3%; Pred. No. 5.8e-92;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
 |||||

Db 1 MSTNPKPQKTKRNTNRRPQDVKFFGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
 |||||

Qy 61 RNNIPKARRDPGRWTWANGYPWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
 |||||

Db 61 RRQIPKARRPEGRTWAQPGYWPVLYNCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 4
US-08-290-665A-159
; Sequence 159, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; FILING DATE: 15-AUG-1994
; APPLICATION NUMBER: US/08/290,665A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 159:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S18
US-08-290-665A-159

Query Match 94.9%; Score 998; DB 2; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSRSPRG 60
Db 1 MSTNPKPQRLTKRNTNRRPQDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSRSPRG 60
QY 61 RRNPIKARPDGRTWANPGYWPVLYNCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
Db 61 RRQIPKARRPEGRTWAQPGYWPVLYNCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191

Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 5
US-08-290-665A-160
; Sequence 160, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; FILING DATE: 15-AUG-1994
; APPLICATION NUMBER: US/08/290,665A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DR4
US-08-290-665A-160

Query Match 94.9%; Score 998; DB 2; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSRSPRG 60
Db 1 MSTNPKPQRLTKRNTNRRPQDVKPPGGQIVGGVYLLPRRGPRLGVATRTKTSRSPRG 60
QY 61 RRNPIKARPDGRTWANPGYWPVLYNCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
Db 61 RRQIPKARRPEGRTWAQPGYWPVLYNCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191

Db 181 LLCLTVPASA 191

RESULT 7

PCT-US95-10398-156

; Sequence 156, Application PC/TUS9510398

; GENERAL INFORMATION:

; APPLICANT: BUKH, J., MILLER, R.H. AND

; APPLICANT: PURCELL, R.H.

; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

; NUMBER OF SEQUENCES: 263

; CORRESPONDENCE ADDRESS:

; ADDRESSER: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10398

; FILING DATE: 15-AUG-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/086,428

; FILING DATE: 29 JUNE 1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/290/665

; FILING DATE: 15 AUGUST 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: RICHARD W. BORK

; REGISTRATION NUMBER: 36,459

; REFERENCE/DOCKET NUMBER: 2026-4116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 156:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 191 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; ORIGINAL SOURCE:

; ORGANISM: homosapiens

; INDIVIDUAL ISOLATE: US11

PCT-US95-10398-156

Query Match 94.9%; Score 998; DB 5; Length 191;

Best Local Similarity 96.3%; Pred. No. 5.8e-92;

Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCNTNRRPQDVKFPGGQIVGGVYLLPRGPRGLGVATRTKTSERSNPRG 60

Db 1 MSTNPKPQRTKNTNRRPQDVKFPGGQIVGGVYLLPRGPRGLGVATRTKTSERSQPRG 60

Qy 61 RRNPDKARPDGRTWANGYPWPLYNECGWAGWLLSPRGSPSGWPTDPRRSNGLG 120

Db 61 RRQIPKARPEGTWAGQYPPWPLYNECGWAGWLLSPRGSPSGWPTDPRRSNGLG 120

Qy 121 KVIDTLTCGFDLGMGYIPLVGAAPLGGARALAHGVRVLEQGVNATGNLPGCSFSIFLLA 180

Db 121 KVIDTLTCGFDLGMGYIPLVGAAPLGGARALAHGVRVLEQGVNATGNLPGCSFSIFLLA 180

Qy 181 LLCLTVPASA 191

Db 181 LLCLTVPASA 191

RESULT 6

US-08-380-160-3

; Sequence 3, Application US/08380160

; Patent No. 6235284

; GENERAL INFORMATION:

; APPLICANT: DALBON, Pascal

; APPLICANT: JOLIVET, Michel

; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE

; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY

; TITLE OF INVENTION: FOR DETECTING THE LATTER

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSER: OLIFF & BERRIDGE

; STREET: P.O. Box 19928

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/380,160

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/057,471

; FILING DATE: 06-MAY-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Berridge, William P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 28682

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-6400

; TELEFAX: (703)836-2787

; TELEX:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 191 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; ORGANISM: Human Hepatitis C Virus

US-08-380-160-3

Query Match 94.9%; Score 998; DB 3; Length 191;

Best Local Similarity 96.3%; Pred. No. 5.8e-92;

Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCNTNRRPQDVKFPGGQIVGGVYLLPRGPRGLGVATRTKTSERSNPRG 60

Db 1 MSTNPKPQRTKNTNRRPQDVKFPGGQIVGGVYLLPRGPRGLGVATRTKTSERSQPRG 60

Qy 61 RRNPDKARPDGRTWANGYPWPLYNECGWAGWLLSPRGSPSGWPTDPRRSNGLG 120

Db 61 RRQIPKARPEGTWAGQYPPWPLYNECGWAGWLLSPRGSPSGWPTDPRRSNGLG 120

Qy 121 KVIDTLTCGFDLGMGYIPLVGAAPLGGARALAHGVRVLEQGVNATGNLPGCSFSIFLLA 180

Db 121 KVIDTLTCGFDLGMGYIPLVGAAPLGGARALAHGVRVLEQGVNATGNLPGCSFSIFLLA 180

Qy 181 LLCLTVPASA 191

Db 181 LLSCLTVPASA 191
|||||

RESULT 8

PCT-US95-10398-157
; Sequence 157, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 157:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S14
PCT-US95-10398-157

Query Match 94.9%; Score 998; DB 5; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
Qy 1 MSTNPKPQRLTNTNRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSNPRG 60
Db 1 MSTNPKPQRTKNTNRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
Qy 61 RRPPIKARPDPGRTWNPYGPWPLYNECGGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
Db 61 RROPPIKARPEGRKTAQPGYPWPLYNECGGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
Qy 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180

Qy 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191
|||||

RESULT 9

PCT-US95-10398-158
; Sequence 158, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: SW1
PCT-US95-10398-158

Query Match 94.9%; Score 998; DB 5; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
Qy 1 MSTNPKPQRLTNTNRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSNPRG 60
Db 1 MSTNPKPQRTKNTNRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
Qy 61 RRPPIKARPDPGRTWNPYGPWPLYNECGGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
Db 61 RROPPIKARPEGRKTAQPGYPWPLYNECGGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
Qy 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180

Wed Aug 25 09:56:36 2004

Qy 181 LLSCLTVPASA 191
| | | | | | | | | |
Db 181 LLSCLTVPASA 191

RESULT 10

PCT-US95-10398-159
; Sequence 159, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 159:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S18
PCT-US95-10398-159

Query Match 94.9%; Score 998; DB 5; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCNTNRPPDVKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSNPRG 60
| | | | | | | | | |
Db 1 MSTNPKPQRTKNTNRPPDVKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
| | | | | | | | | |
Qy 61 RRNPICKARPDPGRTWNPYPMPLVYNEGCGWAGWLLSPRGSPSPSGWGTDPWRRSRNLG 120
| | | | | | | | | |
Db 61 RRQIPKARPPEGTWAPQPGYPWPLYNEGCGWAGWLLSPRGSPSPSGWGTDPWRRSRNLG 120
| | | | | | | | | |
Qy 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVLEGDGVNATGNLPGCSFSIFLLA 180
| | | | | | | | | |

Db 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVLEGDGVNATGNLPGCSFSIFLLA 180
| | | | | | | | | |
Qy 181 LLSCLTVPASA 191
| | | | | | | | | |
Db 181 LLSCLTVPASA 191

RESULT 11

PCT-US95-10398-160
; Sequence 160, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DR4
PCT-US95-10398-160

Query Match 94.9%; Score 998; DB 5; Length 191;
Best Local Similarity 96.3%; Pred. No. 5.8e-92;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCNTNRPPDVKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSNPRG 60
| | | | | | | | | |
Db 1 MSTNPKPQRTKNTNRPPDVKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
| | | | | | | | | |
Qy 61 RRNPICKARPDPGRTWNPYPMPLVYNEGCGWAGWLLSPRGSPSPSGWGTDPWRRSRNLG 120
| | | | | | | | | |
Db 61 RRQIPKARPPEGTWAPQPGYPWPLYNEGCGWAGWLLSPRGSPSPSGWGTDPWRRSRNLG 120
| | | | | | | | | |
Qy 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAARALAHGVRLVLEGDGVNATGNLPGCSFSIFLLA 180
| | | | | | | | | |

Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 12
US-08-635-886C-179
; Sequence 179, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-179

Query Match 94.9%; Score 998; DB 4; Length 450;
Best Local Similarity 96.3%; Pred. No. 1.7e-91;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRGLGVRATKTSERSNPRG 60
Db 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRGLGVRATKTSERSQPRG 60
QY 61 RRPPIKARRPDGRTWANPGYWPPLYGNEGCWAGWLLSPRGRSPSWGPTDPWRRSRNLG 120
Db 61 RRQPIKARRPEGRRTWAQPGYWPPLYGNEGCWAGWLLSPRGRSPSWGPTDPWRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 13
US-08-635-886C-180
; Sequence 180, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 180
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-180
Query Match 94.9%; Score 998; DB 4; Length 450;
Best Local Similarity 96.3%; Pred. No. 1.7e-91;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRGLGVRATKTSERSNPRG 60
Db 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRGLGVRATKTSERSQPRG 60
QY 61 RRPPIKARRPDGRTWANPGYWPPLYGNEGCWAGWLLSPRGRSPSWGPTDPWRRSRNLG 120
Db 61 RRQPIKARRPEGRRTWAQPGYWPPLYGNEGCWAGWLLSPRGRSPSWGPTDPWRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 14
US-08-974-690C-179
; Sequence 179, Application US/08974690C
; Patent No. 6613333
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-179

Query Match 94.9%; Score 998; DB 4; Length 450;
Best Local Similarity 96.3%; Pred. No. 1.7e-91;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRGLGVRATKTSERSNPRG 60
Db 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRGLGVRATKTSERSQPRG 60
QY 61 RRPPIKARRPDGRTWANPGYWPPLYGNEGCWAGWLLSPRGRSPSWGPTDPWRRSRNLG 120
Db 61 RRQPIKARRPEGRRTWAQPGYWPPLYGNEGCWAGWLLSPRGRSPSWGPTDPWRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

Wed Aug 25 09:56:36 2004

```
RESULT 15
US-08-974-690C-180
; Sequence 180, Application US/08974690C
; Patent No. 6613333
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPIOTOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 180
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-180

Query Match          94.9%; Score 998; DB 4; Length 450;
Best Local Similarity 96.3%; Pred. No. 1.7e-91;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      1  MSTNPKPQLTCRNTNRRPQDVKFGGGQIVGGVYLLPRRGPRLGVRATKTSERSNPRG 60
Db      1  MSTNPKPQKTKENTNRRPQDVKFGGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60

QY      61  RNPPIPKARRPDGRTWANPGYPWPLYNCGCGWAGWLLSPRGSRPSPWGPTDPWRRSNLG 120
Db      61  RQPIPKARRPEGRTWAQPGYPWPLYNCGCGWAGWLLSPRGSRPSPWGPTDPWRRSNLG 120

QY      121  KVDTLTTCGFADLMGYIPLVGAPLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180
Db      121  KVDTLTTCGFADLMGYIPLVGAPLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180

QY      181  LLSCLTVPASA 191
Db      181  LLSCLTVPASA 191
```

Search completed: August 25, 2004, 04:29:04
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 04:10:51 ; Search time 39 Seconds
(without alignments)
471.092 Million cell updates/sec

Title: US-09-713-687A-1
Perfect score: 1052
Sequence: 1 MSTNPKPQRLTCRNTNRRPQ.....CSFSIFLLALLSCLTVPASA 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	998	94.9	3011	1	GNWVCH genome polyprotein
2	993	94.4	640	2	JQ1584 genome polyprotein
3	990	94.1	3011	1	GNWVC3 genome polyprotein
4	987	93.8	3010	1	GNWVCU genome polyprotein
5	985	93.6	3011	1	S40770 genome polyprotein
6	982	93.3	441	2	S12707 genome polyprotein
7	981	93.3	189	2	S32740 polyprotein - hepa
8	978	93.0	782	2	S18031 genome polyprotein
9	978	93.0	787	2	PN0677 hypothetical prote
10	978	93.0	3010	1	GNWVTC genome polyprotein
11	974	92.6	513	2	FC1284 genome polyprotein
12	969	92.1	369	2	S21471 genome polyprotein
13	967	91.9	782	2	S19876 genome polyprotein
14	965	91.7	3010	1	A45573 genome polyprotein
15	962	91.4	513	2	A44150 structural protein
16	962	91.4	782	2	S19875 genome polyprotein
17	961	91.3	550	2	JH0711 genome polyprotein
18	960	91.3	520	2	JQ1925 polyprotein - hepa
19	960	91.3	523	2	JQ1926 polyprotein - hepa
20	959	91.2	3010	1	S18030 genome polyprotein
21	952	90.5	3010	1	GNWVTW genome polyprotein
22	944	89.7	874	2	JQ0883 genome polyprotein
23	942	89.5	782	2	S18032 genome polyprotein
24	942	89.5	876	2	PC2219 genome polyprotein
25	939	89.3	3033	1	GNWVJ8 polypeptide - hepa
26	928	88.2	178	2	PS0398 genome polyprotein
27	927	88.1	322	2	JN0265 genome polyprotein
28	926	88.0	874	2	JQ0881 genome polyprotein
29	926	88.0	3033	1	JQ1303 genome polyprotein

ALIGNMENTS

RESULT 1

GNWVCH

genome polyprotein - hepatitis C virus (strain H)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

A>Note: host Homo sapiens (man)

C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001

C:Accession: A36814; A41546

R:Inchauspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Submitted to GenBank, July 1992

A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c

A:Reference number: A36814

A:Accession: A36814

A:Molecule type: genomic RNA

A:Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:g329737; PID:AAA45534.1; PID:g329738

R:Inchauspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991

A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari

A:Reference number: A41546; MUID:92052256; PMID:1658800

A:Contents: annotation

A>Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1319/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-2011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,234

Query Match

Best Local Similarity 94.9%; Score 998; DB 1; Length 3011;

Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKEPGGQIVGGVYLLPRGPRGLGVTRATKTSERSNPRG 60

Db 1 MSTNPKPQRLTCRNTNRRPQDVKEPGGQIVGGVYLLPRGPRGLGVTRATKTSERSNPRG 60

QY 61 RRNPPIKARPDGRITWANPGYWPPLYGNECGWAGWLLSPRGSRPSPMGTPDWRSRNLG 120

Db 61 RRNPPIKARPDGRITWANPGYWPPLYGNECGWAGWLLSPRGSRPSPMGTPDWRSRNLG 120

QY 121 KVITDTTCGFADLMGYPILVGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPLVGLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180
181 LLSCLTVPASA 191
181 LLSCLTVPASA 191

RESULT 2
JQ1584
genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
A:Accession: JQ1584
R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative core protein NS4a
A:Reference number: JQ1584; MUID:92300349; PMID:1318944
A:Accession: JQ1584
A:Molecule type: genomic RNA
A:Residues: 1-640 <KUM>
A:Cross-references: GB:X84079; NID:g643119; PIDN:CAAS8898.1; PID:g643120
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polyprotein
F:1-191/Product: core protein C #status predicted <CPC>
F:192-389/Product: envelope protein E1 #status predicted <E1>
F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <NS1>
F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (coverage)
Query Match 94.4%; Score 993; DB 2; Length 640;
Best Local Similarity 95.8%; Pred. No. 1.4e-77;
Matches 183; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCENTNRPDQVKFPGGGQIVGGVLLPRGPRLGVTRATKTSERNPRG 60
Db 1 MSTNPKPQRLTCENTNRPDQVKFPGGGQIVGGVLLPRGPRLGVTRATKTSERNPRG 60

QY 61 RNPPIKARRPDORTWANPGYPWLYNCGCGWAGLLSPGRSPSGWPTDPMRRSRNLG 120
Db 61 RQPIKARRPEGRITWAQPGYPWLYNCGCGWAGLLSPGRSPSGWPTDPMRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 3
GNMVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
A:Accession: A39166; MUID:91172826; PMID:1848704
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coombs, R.W.; Omish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.; Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to other hepatitis C virus types
A:Reference number: P00403; MUID:92268871; PMID:1316939
A:Accession: P00403
A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: P00404
A:Status: Preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
F:116-191/Product: envelope protein C #status predicted <CPC>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,2222
Query Match 94.1%; Score 990; DB 1; Length 3011;
Best Local Similarity 95.3%; Pred. No. 1.1e-76;
Matches 182; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCENTNRPDQVKFPGGGQIVGGVLLPRGPRLGVTRATKTSERNPRG 60
Db 1 MSTNPKPQRLTCENTNRPDQVKFPGGGQIVGGVLLPRGPRLGVTRATKTSERNPRG 60

QY 61 RNPPIKARRPDORTWANPGYPWLYNCGCGWAGLLSPGRSPSGWPTDPMRRSRNLG 120
Db 61 RQPIKARRPEGRITWAQPGYPWLYNCGCGWAGLLSPGRSPSGWPTDPMRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGLGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 4
GNMVC3
genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
A:Accession: A39253; PMID:2175903
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, K.; Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variation
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KAT>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine protease
F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEB>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 93.8%; Score 987; DB 1; Length 3010;
Best Local Similarity 95.3%; Pred. No. 28-76;
Matches 182; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKFPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSNPRG 60
DB 1 MSTNPKPQRTKRNTRRRPQDVKFPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60

QY 61 RRNPPIKARRPDGRTWANGPYWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
DB 61 RRQPIPKARRPEGRWTAQPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
DB 181 LLSCLTIPASA 191

RESULT 5
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEB>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 93.6%; Score 985; DB 1; Length 3011;
Best Local Similarity 95.3%; Pred. No. 3e-76;
Matches 182; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKFPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSNPRG 60
DB 1 MSTNPKPQRTKRNTRRRPQDVKFPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60

QY 61 RRNPPIKARRPDGRTWANGPYWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
DB 61 RRQPIPKARRPEGRWTAQPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
DB 181 LLSCLTIPASA 191

RESULT 6
S12707
genome polyprotein - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: S12707
R:Takeuchi, K.; Kubo, Y.; Boonmar, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G.; Nucleic Acids Res. 18, 4626, 1990
A:Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus genome
A:Reference number: S12707; MUID:90356432; PMID:2117749
A:Accession: S12707
A:Molecule type: genomic RNA
A:Residues: 1-441 <TAK>
A:Cross-references: EMBL:D00574; NID:g221656; PIDN:BAA00452.1; PID:g221657
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 93.3%; Score 982; DB 2; Length 441;
Best Local Similarity 94.8%; Pred. No. 8.5e-77;
Matches 181; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKFPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSNPRG 60
DB 1 MSTNPKPQRTKRNTRRRPQDVKFPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60

QY 61 RRNPPIKARRPDGRTWANGPYWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120
DB 61 RRQPIPKARRPEGRWTAQPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPWRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAFLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
DB 181 LLSCLTIPASA 191

RESULT 7
S32740
polyprotein - hepatitis C virus (isolate Russian) (fragment)
N:Contains: capsid protein C; envelope protein M
C:Species: hepatitis C virus
C:Date: 19-Mar-1997 #sequence_revision 05-Dec-1998 #text_change 17-Nov-2000
C:Accession: S32740
R:Vassiliev, V.B.; Viazov, S.O.; Kotova, E.Y.; Nosikov, V.V.
submitted to the EMBL Data Library, April 1993
A:Description: Evidence of new HCV variant of European isolate in Russia.
A:Reference number: S32740
A:Accession: S32740
A:Molecule type: genomic RNA

A;Residues: 1-189 <VAS>

A;Cross-references: EMBL:X71407
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; envelope protein; polyprotein
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-189/Product: envelope protein M #status predicted <EPM>

Query Match 93.3%; Score 981; DB 2; Length 189;
Best Local Similarity 95.8%; Pred. No. 4.6e-77;
Matches 181; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPDVKFPGGQIYGGVYLLPRGPRGLGVRATKTSERSNPRG 60
DB 1 MSTNPKPQRTKXNTNRRPDVKFPGVQIYGGVYLLPRGPRGLGVRATKTSERSQPRG 60

QY 61 RRNPDKARRPDGRTWANPGYPMPLYNEGCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
DB 61 RRQPIKARRPEGRRTWAQPGYPMPLYNEGCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPA 189
DB 181 LLSCLTTPA 189

RESULT 8

S18031
genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JK2
C;Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text change 17-Nov-2000
C;Accession: S18031
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A;Reference number: S18029
A;Accession: S18031
A;Molecule type: genomic RNA
A;Residues: 1-782 <HON>
A;Cross-references: EMBL:X61593
A;Experimental source: isolate JK2
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <M1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 93.0%; Score 978; DB 2; Length 782;
Best Local Similarity 94.2%; Pred. No. 3.2e-76;
Matches 180; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPDVKFPGGQIYGGVYLLPRGPRGLGVRATKTSERSNPRG 60
DB 1 MSTNPKPQRTKXNTNRRPDVKFPGGQIYGGVYLLPRGPRGLGVRATKTSERSQPRG 60

QY 61 RRNPDKARRPDGRTWANPGYPMPLYNEGCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120
DB 61 RRQPIKARRPEGRRTWAQPGYPMPLYNEGCGWAGWLLSPRGSRPSWGPTDPRRSRNIG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPA 191
DB 181 LLSCLTTPA 191

RESULT 9

PN0677
hypothetical protein 787 - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PN0677
R;Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.
Biochem. Biophys. Res. Commun. 196, 780-788, 1993
A>Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of genotypes and their association with liver disease
A:Reference number: PN0677; MUID:94059104; PMID:8240354
A:Accession: PN0677
A:Molecule type: mRNA
A:Residues: 1-787 <CHO>
A:Cross-references: GB:L20498; NID:gl381031; PIDN:AAB02608.1; PID:gl381032
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; nonstructural protein
F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate binding site
Query Match 93.0%; Score 978; DB 2; Length 787;
Best Local Similarity 94.2%; Pred. No. 3.3e+76;
Matches 180; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MSTNPQRQLTCNTNRPPDVKFPGGQIVGVYLLPRGPGPLGLVRAATRKTSERNPRG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MSTNPQRKTENRRPDDVKFPFGQIVGVYLLPRGPGPLGLVRAATRKTSERSQPRG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 61 RRPIPKARPDGRWTANPCGYPMPLVGNCEGWAGWLLSPRGSRPSWGPTDPRWRRNLG 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 RRPIPKARPEGRAWAQPGYPWLTVNEGLGWAGWLLSPRGSRPSWGPTDPRRSNLG 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAALAHGVRVLEDGVNYATGNLPGCFSIFLLA 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAAALAHGVRVLEDGVNYATGNLPGCFSIFLLA 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 181 LLSCLITVPASA 191
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 LLSCLITIPASA 191
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10

GNWVC
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991
A>Title: Structure and organization of the hepatitis C virus genome isolated from human liver tissue
A:Reference number: A38465; MUID:91140698; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TAK>
A:Cross-references: EMBL:M59335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein
F:12-115/Product: capsid protein C #status predicted <CP>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (p-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2
Query Match 93.0%; Score 978; DB 1; Length 3010;
Best Local Similarity 94.8%; Pred. No. 1.2e-75;

Matches 181; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Db 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Qy 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Db 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Qy 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTVPASA 191
 Db 181 LLSCLTVPASA 191

RESULT 11
 PC1284
 genome polyprotein - hepatitis C virus (isolate HC-J4) (fragment)
 C:Species: hepatitis C virus
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: PC1284; PMID:2170712
 A:Accession: PC1284
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OKA>
 A:Cross-references: GB:D00832; NID:g221513; PIDN:BAA00706.1; PID:g221514
 C:Superfamily: hepatitis C virus genome polyprotein

Query Match 92.6%; Score 974; DB 2; Length 513;
 Best Local Similarity 94.2%; Pred. No. 4.8e-76;
 Matches 180; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Db 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Qy 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Db 61 WRQPIKARRPEGRAWAQPWPPLYGNEGGLGAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Qy 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTVPASA 191
 Db 181 LLSCLTVPASA 191

RESULT 12
 S21471
 genome polyprotein - hepatitis C virus (fragment)
 N:Contains: capsid protein; envelope protein
 C:Species: hepatitis C virus
 C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
 C:Accession: S21471
 R:Mogan, W.K.
 submitted to the EMBL Data Library, April 1992
 A:Reference number: S21471
 A:Accession: S21471
 A:Molecule type: genomic RNA
 A:Residues: 1-369 <MOG>
 A:Cross-references: EMBL:X65924; NID:g59466; PIDN:CAA46717.1; PID:g59467
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein

Query Match 92.1%; Score 969; DB 2; Length 369;
 Best Local Similarity 93.7%; Pred. No. 9.3e-76;
 Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Db 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Qy 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Db 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Qy 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTVPASA 191
 Db 181 LLSCLTVPASA 191

RESULT 13
 S19876
 genome polyprotein - hepatitis C virus (isolate JKS) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JKS
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 R:Honda M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A:Reference number: S18029
 A:Accession: S19876
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
 A:Experimental source: isolate JKS
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F:1-191/Product: core protein #status predicted <MAT1>
 F:192-363/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 91.9%; Score 967; DB 2; Length 782;
 Best Local Similarity 93.2%; Pred. No. 2.9e-75;
 Matches 178; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Db 1 MSTNPKPQRLTCRNTNRRPDQVKFPGGQIVGGVYLLPRRGRLGVRAATKTSERSNPRG 60
 Qy 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Db 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGRSPSWGPTDPWRSRNLG 120
 Qy 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTVPASA 191
 Db 181 LLSCLTVPASA 191

RESULT 14
 A45573
 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: A45573		Rifanaka, I.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Y.		61		RRQPIPKARRPEGRRTWAQFGYPWPPLYGNEGCGWAGWLLSPRGSRGFSWG-TDPRRRSRNL		119	
Virus Res. 23, 39-53, 1992									
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s				QY		120		GKVIDTLTCGFADLMGYIPLVGCAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLL	
A:Reference number: A45573; PMID:92295714; PMID:1318627				Db		120		GKVIDTLTCGFADLMGYIPLVGCAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLL	
A:Accession: A45573				QY		180		ALLSCLTVPASA 191	
A:Status: preliminary				Db		180		ALLSCLTVPASA 191	
A:Molecule type: DNA									
A:Residues: 1-3010 <TAN>									
A:Cross-references: GB:D11168; GB:D01171; NID:G221612; PIDN:BAA01943.1; PID:G221613									
A:Experimental source: HCV-JT									
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)									
C:Superfamily: hepatitis C virus genome polyprotein									
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin									
F:2-115/Product: capsid protein C #status predicted <CPC>									
F:116-191/Product: envelope protein M #status predicted <EPM>									
F:192-389/Product: major envelope protein E #status predicted <MEE>									
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>									
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>									
F:1007-1615/Product: hepatitis virus #status predicted <NS3>									
F:1230-1237/Region: nucleotide-binding motif A (P-loop)									
F:1312-1317/Region: nucleotide-binding motif B									
F:1316-1319/Region: DEXH motif									
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>									
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>									
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>									
Query Match		91.7%; Score 965; DB 1; Length 3010;							
Best Local Similarity		93.2%; Pred. No. 1.6e-74;							
Matches		178; Conservative 3; Mismatches 10; Indels 0; Gaps 0;							
QY		1		MSTNPKPQRLTCRNTNRPPQDVKFGGQIVGGVYLLPRRGLGVRAATKTSERSNPRG		60			
Db		1		MSTNPKPQKTKRNTYRRPDQVKKFGGQIVGGVYLLPRRGLGVRAATKTSERSQPRG		60			
QY		61		RRNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLG		120			
Db		61		RRQPIPKARRPEGRAWAQFGYPWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLG		120			
QY		121		KVIDTLTCGFADLMGYIPLVGCAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA		180			
Db		121		KVIDTLTCGFADLMGYIPLVGCAPLGGARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA		180			
QY		181		LLSCLTVPASA 191					
Db		181		LLSCLTVPASA 191					
RESULT 15									
A44150									
structural protein - hepatitis C virus									
C:Species: hepatitis C virus									
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000									
C:Accession: A44150									
R:Ching, W.M.; Wychoowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.									
Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992									
A:Title: Interaction of immune sera with synthetic peptides corresponding to the structu									
A:Reference number: A44150; PMID:92228749; PMID:1373489									
A:Accession: A44150									
A:Status: preliminary; not compared with conceptual translation									
A:Molecule type: genomic RNA									
A:Residues: 1-513 <CHI>									
C:Superfamily: hepatitis C virus genome polyprotein									
Query Match		91.4%; Score 962; DB 2; Length 513;							
Best Local Similarity		94.8%; Pred. No. 5.1e-75;							
Matches		182; Conservative 1; Mismatches 7; Indels 2; Gaps 2;							
QY		1		MSTNPKPQRLTCRNTNRPPQDVKFGGQIVGGVYLLPRRGLGVRAATKTSERSNPRG		60			
Db		1		MSTNPKPQIKTKRNTNRPPQDVKFGGQIVGGVYLLPRRGLGVRAATKTSERSQPRG		60			
QY		61		RRNPPIKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGSR-PSWGPTDPRRRSRNL		119			

Search completed: August 25, 2004, 04:28:26
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 04:02:45 ; Search time 23 Seconds

(without alignments)
432.409 Million cell updates/sec

Title: US-09-713-687A-1

Perfect score: 1052

Sequence: 1 MSTNPKPQRLTCRNTNRRPQ.....CSFSIFLLALLSCLTVTPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	998	94.9	3011	1	POLG_HCVH
2	990	94.1	3011	1	POLG_HCVI
3	987	93.8	3010	1	POLG_HCVJA
4	978	93.0	3010	1	POLG_HCVBK
5	976	92.8	513	1	POLG_HCVJ2
6	965	91.7	3010	1	POLG_HCVJ7
7	960	91.3	520	1	POLG_HCVH4
8	960	91.3	520	1	POLG_HCVHK
9	952	90.5	3010	1	POLG_HCVTW
10	944	89.7	737	1	POLG_HCVJ7
11	939	89.3	3033	1	POLG_HCVJ8
12	934	88.8	737	1	POLG_HCVJ5
13	926	88.0	3033	1	POLG_HCVJ6
14	385	36.6	192	1	POLG_HCVJ1
15	385	36.6	321	1	POLG_HCVH8
16	383	36.4	321	1	POLG_HCVTH
17	322	30.6	309	1	POLG_HCVH7
18	83.5	7.9	443	1	SPA4_MOUSE
19	83.5	7.9	512	1	IE63_HSV11
20	82.5	7.8	252	1	PRIO_RABIT
21	82.5	7.8	444	1	SPA4_RAT
22	82	7.7	382	1	METX_MYCLE
23	81.5	7.7	348	1	YT35_STRFR
24	81	7.7	1321	1	IRS2_MOUSE
25	80	7.6	257	1	PRIO_MUSPF
26	80	7.6	257	1	PRIO_MUSVI
27	80	7.6	566	1	PKNE_MYCTU
28	79.5	7.6	256	1	PRIO_CAPH1
29	79.5	7.6	256	1	PRIO_FELCA
30	79.5	7.6	256	1	PRIO_SHEEP
31	79.5	7.6	256	1	PRP2_BOVIN
32	79	7.5	628	1	V70K_TYMW
33	78.5	7.5	1838	1	CA15_HUMAN

34	77.5	7.4	255	1	PRIO_CANFA
35	77	7.3	602	1	RHO_MYCTU
36	77	7.3	750	1	ELS_CHICK
37	76.5	7.3	256	1	PRIO_CEREL
38	76.5	7.3	674	1	CA1A_CHICK
39	76	7.2	461	1	FALC_CHRVO
40	76	7.2	560	1	YELL_DROPS
41	76	7.2	4829	1	BIR6_HUMAN
42	75.5	7.2	252	1	DUT_HUMAN
43	75.5	7.2	342	1	ROAI_SCHAM
44	75.5	7.2	437	1	PROP_MOUSE
45	75.5	7.2	551	1	YOLI_MYCTU

O46501 canis famil
Q10607 mycobacteri
P07916 gallus gall
P79142 cervus elap
P08125 gallus gall
Q7N202 chromobacte
Q9B118 drosophila
Q9N709 homo sapien
P33316 homo sapien
P21522 schistocerc
P11680 mus musculu
P71732 mycobacteri

RESULT 1

POLG_HCVH	STANDARD;	PRT; 3011 AA.
AC	P27958;	
DT	01-AUG-1992 (Rel. 23, Created)	
DT	01-AUG-1992 (Rel. 23, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Genome polypeptide [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].	
OS	Hepatitis C virus (isolate H) (HCV).	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.	
OX	NCBI_TaxID=11108;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=92052256; PubMed=1658800;	
RA	Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M., Prince A.M.;	
RA	"Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.";	
RT	Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).	
RN	[2]	
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.	
RX	MEDLINE=97331322; PubMed=9187654;	
RA	Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;	
RT	"Structure of the hepatitis C virus RNA helicase domain.";	
RL	Nat. Struct. Biol. 4:463-467(1997).	
RN	[3]	
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.	
RX	MEDLINE=98154321; PubMed=9493270;	
RA	Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R.;	
RT	"Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.";	
RL	Structure 6:89-100(1998).	
CC	- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.	
CC	- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.	
CC	- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.	
CC	- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.	
CC	- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.	
CC	- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.	
CC	- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).	
CC	- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1	

ALIGNMENTS

CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M67463; AAA45334.1; -;
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR TRANSFAC; T04155; -;
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS4b.
DR InterPro; IPR002166; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR004109; Peptidase C29.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 746
FT CHAIN 747 809
FT CHAIN 810 1026
FT CHAIN 1027 1657
FT CHAIN 1658 1711
FT CHAIN 1712 1972
FT CHAIN 1973 2420
FT CHAIN 2421 3011
FT CHAIN 3011 369
FT TRANSMEM 347 1083
FT ACT_SITE 1083 1107
FT ACT_SITE 1107 1165
FT ACT_SITE 1165 1237
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196

1029 209 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
234 234 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
305 305 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
417 417 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
423 423 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
430 430 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
448 448 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
476 476 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
532 532 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
540 540 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
556 556 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
576 576 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
623 623 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
645 645 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1224 1224 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1232 1232 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1238 1238 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1246 1246 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1248 1248 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1251 1251 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1255 1255 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1271 1271 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1272 1272 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1277 1277 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1281 1281 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1283 1283 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1291 1291 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1295 1295 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1301 1301 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1302 1302 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1312 1312 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1317 1317 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1323 1323 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1335 1335 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1343 1343 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1352 1352 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1360 1360 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1362 1362 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1368 1368 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1373 1373 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1376 1376 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1378 1378 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1382 1382 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1389 1389 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1397 1397 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1410 1410 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1411 1411 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1414 1414 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1419 1419 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1432 1432 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1438 1438 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1450 1450 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1456 1456 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1471 1471 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1480 1480 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1481 1481 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1489 1489 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1497 1497 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1501 1501 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1511 1511 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1514 1514 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1532 1532 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1550 1550 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1555 1555 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1570 1570 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1579 1579 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1580 1580 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1597 1597 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1598 1598 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1606 1606 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1611 1611 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1622 1622 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1627 1627 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1635 1635 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1640 1640 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
1652 1652 CARBOHYD N-LINKED (GLCNAC. .) (POTENTIAL).
3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
SEQUENCE

Query Match 94.9%; Score 998; DB 1; Length 3011;
 Best Local Similarity 96.3%; Pred. No. 6.5e-78;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCNTNRRPDQVFPGGQIVGGVYLLPRGPRGLGVATKTKTSERNPRG 60
 DB 1 MSTNPKPQRLTCNTNRRPDQVFPGGQIVGGVYLLPRGPRGLGVATKTKTSERQPRG 60
 QY 61 RRPPIKARRPDGRTWANPGYWPVLYGNECGWAGLLSPRGSRPSMGPTDPRRSRNLG 120
 DB 61 RRPPIKARRPDGRTWANPGYWPVLYGNECGWAGLLSPRGSRPSMGPTDPRRSRNLG 120
 QY 121 KVIDLTCGFADLMGVIPLVGPALGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180
 DB 121 KVIDLTCGFADLMGVIPLVGPALGGAARALAHGVRLDGVNYATGNLPGCSFSIFLLA 180
 QY 181 LLSCLTVPASA 191
 DB 181 LLSCLTVPASA 191

RESULT 2
 POLG HCV1
 ID POLG HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OK NCBI_Taxid=11104;
 [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M62321; AAA5676.1;
 CC PIR; A39166; GNWVC3.
 CC PDB; 1A1V; 16-FEB-99.
 CC PDB; 1HEI; 25-NOV-98.

DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_psvir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 DR 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1515 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1516 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 POTENTIAL.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCB5AF9 CRC64;

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

Query Match	94.1%;	Score 990;	DB 1;	Length 3011.
Best Local Similarity	95.3%;	Pred. No. 3.2e-77;		
Matches 182; Conservative	2;	Mismatches 7;	Indels 0;	Gaps 0;
<hr/>				
QY	1	MSNPKPQRLTCNTNRPPDVFPGGGQIVGGVYLPRRCPRLGVRAATKTSRRSPRG	60	
Db	1	MSNPKPQKNKNTNRPPDVFPGGGQIVGGVYLPRRCPRLGVRAATKTSRSQDRG	60	
<hr/>				
QY	61	RRPIPKARPPDGRTWANPGYPFLYNGECGWAGLLSPGRSPSWGPTDPWRRSRNLG	120	
Db	61	RROPIKARRPEGRTWAPGYEPLYNGECGWAGLLSPGRSPSWGPTDPERRSRNLG	120	
<hr/>				
QY	121	KVIDTLTCGFADLMGYIPLVGAPLGGGAARALAHGVRVLEDGVYATGNLPGCSFSIFLLA	180	
Db	121	KVIDTLTCGFADLMGYIPLVGAPLGGGAARALAHGVRVLEDGVYATGNLPGCSFSIFLLA	180	
<hr/>				
QY	181	LLSCLTVPASA	191	
Db	181	LLSCLTVPASA	191	

RESULT 3

POLG	HCVJA	STANDARD;	PRT;	3010 AA.
ID	POLG HCVJA			
AC	P26662;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);			
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2			
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)			
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)			
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein			
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein			
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate Japanese) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11116;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=91088550; PubMed=2175903;			
RX	Kato N., Hijikata M., Nakagawa Y., Nakagawa M., Ohkoshi S.,			
RA	Sugimura T., Shimotohno K.;			
RA	"Molecular cloning of the human hepatitis C virus genome from			
RT	Japanese patients with non-A, non-B hepatitis.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).			
[2]				
RN	DISCUSSION OF SEQUENCE.			
RP	MEDLINE=91192160; PubMed=1849488;			
RX	Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,			
RA	Ohkoshi S., Shimotohno K.;			
RA	"Molecular structure of the Japanese hepatitis C viral genome.;"			
RL	FEBS Lett. 280:325-328(1991).			
CC	-!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are			
CC	hydrophobic, suggesting a possible membrane-related function. NS3			
CC	and NS5 may play a role in the viral RNA replication.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral			
CC	precursor polyprotein, commonly with Asp or Glu in the P6			
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +			
CC	{N} (N).			
CC	-!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a			
CC	lipoprotein envelope. The envelope consists of two proteins:			
CC	protein M and glycoprotein E. The nucleocapsid is a complex of			
CC	protein C and mRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			

FT	CHAIN	1616	1862	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).	
FT	CHAIN	1863	2013	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).	
FT	CHAIN	2014	3010	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).	
FT	TRANSMEM	347	369	POTENTIAL.	
FT	ACT_SITE	1083	1083	CHARGE RELAY SYSTEM.	
FT	ACT_SITE	1107	1107	CHARGE RELAY SYSTEM.	
FT	ACT_SITE	1165	1165	CHARGE RELAY SYSTEM.	
FT	NP_BIND	1230	1237	ATP (POTENTIAL).	
FT	SITE	1316	1319	DECH BOX.	
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	2041	2041	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	2077	2077	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	2240	2240	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	2529	2529	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	2788	2788	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1031	1035	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	HELIX	1039	1047	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1050	1050	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1059	1063	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1068	1074	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1075	1076	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1077	1081	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	HELIX	1082	1085	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1086	1087	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1090	1092	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1093	1094	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1095	1097	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1101	1103	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1104	1107	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1108	1112	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1120	1120	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1122	1122	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1129	1133	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1135	1136	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1139	1144	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1149	1157	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	HELIX	1158	1161	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1162	1163	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1165	1166	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1168	1171	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1172	1174	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1175	1186	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1187	1188	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1189	1197	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	HELIX	1198	1202	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	TURN	1203	1204	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	STRAND	1680	1688	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SQ	SEQUENCE	3010 AA;	327189 MW;	F8422D5ECCCFDFD9C CRC64;	
Query Match 93.0%; Score 978; DB 1; Length 3010;					
Best Local Similarity 94.8%; Pred. No. 3.4e-76;					
Matches 181; Conservative 1; Mismatches 9; Indels 0; Gaps 0;					
Qy	1	MSINPKPQRLTCRNTNRRPDQVKPFGGQIVGGVYLLPRRGPRLGVRATRTSRSNPRG	60		
Db	1	MSINPKPQRTKRNINRRPDQVKPFGGQIVGGVYLLPRRGPRLGVRAPRTSRSQPRG	60		
Qy	61	RRNPPIKARRPDGRTWANGVYWPVLYGNEGCGWALLSPRGSRPSWGPTDPRRSRLG	120		

Db

61

RRQPIKARRPDGRTWANGVYWPVLYGNEGCGWALLSPRGSRPSWGPTDPRRSRLG

120

Qy

121

KVIDTLTCGFADLMGYIPLVNGAPLGGAARALAHGVRLVEDGVNATGNLPGCSFIFILIA

180

Db

121

KVIDTLTCGFADLMGYIPLVNGAPLGGAARALAHGVRLVEDGVNATGNLPGCSFIFILIA

180

Qy

181

LLSCLTVPASA

191

Db

181

LLSCLTTPASA

191

RESULT 5

POLG HCVJ2

STANDARD;

PRT;

513 AA.

ID

POLG HCVJ2

AC

P27959;

DT

01-AUG-1992

(Rel. 23, Created)

DT

01-AUG-1992

(Rel. 23, Last sequence update)

DT

16-OCT-2001

(Rel. 40, Last annotation update)

DE

Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE

Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE

(GP68) (GP70) (NS1)] (Fragment).

OS

Hepatitis C virus (isolate HC-J2) (HCV).

OC

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC

Hepacivirus.

OX

NCBI_TaxID=11111;

[1]

RP

SEQUENCE FROM N.A.

RA

MEDLINE=92230232; PubMed=1314459;

RA

Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,

RA

Tanaka T., Fukuda S., Tsuda F., Mishiro S.;

RT

"Full-length sequence of a hepatitis C virus genome having poor

RT

homology to reported isolates: comparative study of four distinct

RT

genotypes."

RL

Virology 188:331-341(1992).

CC

-!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are

CC

hydrophobic, suggesting a possible membrane-related function. NS3

CC

and NS5 may play a role in the viral RNA replication.

CC

-!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a

CC

lipoprotein envelope. The envelope consists of two proteins:

CC

protein M and glycoprotein E. The nucleocapsid is a complex of

CC

protein C and mRNA.

CC

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation

CC

the European Bioinformatics Institute. There are no restrictions on its

CC

use by non-profit institutions as long as its content is in no way

CC

modified and this statement is not removed. Usage by and for commercial

CC

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC

or send an email to license@isb-sib.ch).

CC

CC

EMBL; D10074; BAA00968.1; -

DR

InterPro; IPR002522; HCV capsid.

DR

InterPro; IPR002521; HCV core.

DR

InterPro; IPR002519; HCV env.

DR

InterPro; IPR002531; HCV NS1.

DR

Pfam; PF01543; HCV capsid; 1.

DR

Pfam; PF01542; HCV core; 1.

DR

Pfam; PF01539; HCV env; 1.

DR

Pfam; PF01560; HCV NS1; 1.

DR

ProDom; PD186062; HCV NS1; 1.

KW

Polyprotein; Glycoprotein; Coat protein; Envelope protein;

KW

Transmembrane; Nonstructural

FT

INIT_MET 1

FT

CHAIN 1 115

FT

CHAIN 116 191

FT

CHAIN 192 383

FT

CHAIN 384 >513

FT

CHAIN 347 369

FT

TRANSMEM 196 196

FT

CARBOHYD 209 209

FT

CARBOHYD 233 233

FT

CARBOHYD 234 234

FT

REMOVED FROM CAPSID PROTEIN C BY THE

FT

CELLULAR AMINOPEPTIDASE.

FT

CAPSID PROTEIN C (POTENTIAL).

FT

MATRIX PROTEIN (POTENTIAL).

FT

MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT

NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).

FT

POTENTIAL.

FT

N-LINKED (GLCNAC. .) (POTENTIAL).

FT

N-LINKED (GLCNAC. .) (POTENTIAL).

FT

N-LINKED (GLCNAC. .) (POTENTIAL).

FT

N-LINKED (GLCNAC. .) (POTENTIAL).


```
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match
Best Local Similarity 91.7%; Score 965; DB 1; Length 3010;
Matches 178; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCNTNRRPDQVKFGGQIVGGVYLLPRGRLGVTRATKTSERSNPRG 60
DB 1 MSTNPKPQKTKRNTNRRPDQVKFGGQIVGGVYLLPRGRLGVTRATKTSERSQPRG 60
QY 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCWAGWLLSPRSGRSPSWGPTDPRRSRNLG 120
DB 61 RRPQIPKARRPEGRAWAQGYWPPLYGNEGLWAGWLLSPRSGRSPSWGPTDPRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
DB 181 LLSCLTIPASA 191

RESULT 7
POLG_HCVH4
ID POLG_HCVH4 STANDARD; PRT; 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment)
OS Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992)
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D10688; BAA01530.1; -
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC
```

```
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520
SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match
Best Local Similarity 92.7%; Score 960; DB 1; Length 520;
Matches 177; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCNTNRRPDQVKFGGQIVGGVYLLPRGRLGVTRATKTSERSNPRG 60
DB 1 MSTNPKPQKTKRNTNRRPDQVKFGGQIVGGVYLLPRGRLGVTRATKTSERSQPRG 60
QY 61 RNPPIKARRPDGRTWANPGYWPPLYGNEGCWAGWLLSPRSGRSPSWGPTDPRRSRNLG 120
DB 61 RRPQIPKARRPEGRAWAQGYWPPLYGNEGLWAGWLLSPRSGRSPSWGPTDPRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
DB 181 LMSCLTAPASA 191

RESULT 8
POLG_HCVHK
ID POLG_HCVHK STANDARD; PRT; 520 AA.
AC Q01403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment)
OS Hepatitis C virus (isolate HCV-KF) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992)
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```


entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC  BMBL; DI0687; BAA01529.1; --
CC  PIR; JQ1925; JQ1925; HCV capsid.
DR  InterPro; IPR002522; HCV capsid.
DR  InterPro; IPR002521; HCV core.
DR  InterPro; IPR002519; HCV env.
DR  InterPro; IPR002531; HCV NS1.
DR  Pfam; PF01543; HCV capsid; 1.
DR  Pfam; PF01542; HCV core; 1.
DR  Pfam; PF01539; HCV env; 1.
DR  Pfam; PF01560; HCV NS1; 1.
DR  ProDom; PD186062; HCV NS1; 1.
KW  Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW  Transmembrane; Nonstructural protein.
FT  INIT_MET 1
FT  CHAIN 1 115
FT  CHAIN 116 191
FT  CHAIN 192 383
FT  CHAIN 384 >520
FT  TRANSMEM
FT  CARBOHYD 196 196
FT  CARBOHYD 209 209
FT  CARBOHYD 233 233
FT  CARBOHYD 234 234
FT  CARBOHYD 305 305
FT  CARBOHYD 418 418
FT  CARBOHYD 424 424
FT  CARBOHYD 431 431
FT  CARBOHYD 449 449
FT  NON TER 520 520
SQ  SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 91.3%; Score 960; DB 1; Length 520;
Best Local Similarity 92.7%; Pred. No. 2.1e-75;
Matches 177; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTRRPQDVKPGGGQIVGGVYLLPRGRIGRIVARTKTSRNPGRG 60
DB 1 MSTNPKPQRLTCRNTRRPQDVKPGGGQIVGGVYLLPRGRIGRIVARTKTSRNPGRG 60
QY 61 RRNPPIKARRPDGRTWANGPYWPLVGNCGWAGLLSPGRSPSWGTPDPWRSRNLG 120
DB 61 RRNPPIKARRPDGRTWANGPYWPLVGNCGWAGLLSPGRSPSWGTPDPWRSRNLG 120
QY 121 KVITLTCGFADLMGYIPLVCAPIGGARALAHGVRLVDGVNATGNLPGCSFSIFLLA 180
DB 121 KVITLTCGFADLMGYIPLVCAPIGGARALAHGVRLVDGVNATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
DB 181 LMSCLTVPASA 191

POLG_HCVTW STANDARD; PRT; 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

```

```

OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RT mapping the 5' termini of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84754; -; NOT ANNOTATED CDS.
DR PIR; A40244; GNVVTV.
DR PDB; 1N64; 25-FEB-03.
DR PDB; 1NS3; 08-APR-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4A.
DR InterPro; IPR001490; HCV NS4B.
DR InterPro; IPR002868; HCV NS5A.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4A; 1.
DR Pfam; PF01001; HCV NS4B; 1.
DR Pfam; PF01506; HCV NS5A; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF000998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383

```



```

QY 121 KVIDTLCGFADLMGYIPVVGAPLGGAAALAHGVRVLEQGVNATGNLPGCCSFSIFLLA 180
Db 121 KVIDTLCGFADLMGYIPVVGAPVGGVARALAHGVRVLEQGVNATGNLPGCCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPVSA 191

RESULT 11
POLG_HCVJ8
ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN [1]_TaxID=11115;
RP SEQUENCE FROM N.A.
RX MEDLINE=92230332; PubMed=1114459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10988; BAA01761.1; -.
DR PIR: A40250; GNMVJ8.
DR HSP: P27958; 1HE1.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV env.
DR InterPro: IPR002519; HCV core.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.

```

```

DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DRXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 89.3%; Score 939; DB 1; Length 3033;
Best Local Similarity 88.5%; Pred. No. 7,9e-73;
Matches 169; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 1 MSTNPKPQLTCRNTNRRPDVFPGGQVGVVLLPRGGPRLGVPATKTSRSPRG 60
Db 1 MSTNPKPQRTKNTNRRPDVFPGGQVGVVLLPRGGPRLGVPATKTSRSPRG 60
QY 61 RRNPPIKARPDPGRGTWANPGYWPFLYGNCGWAGWLLSPRGRSPSGPTDPPWRSNRLG 120
Db 61 RRQPIPKDRESTGKSGKPGYWPFLYGNCGWAGWLLSPRGRSPSGPTDPPHRSNRLG 120
QY 121 KVIDTLCGFADLMGYIPVVGAPLGGAAALAHGVRVLEQGVNATGNLPGCCSFSIFLLA 180
Db 121 RVIDTTCGFADLMGYIPVVGAPVGGVARALAHGVRVLEQGVNATGNLPGCCSFSIFLLA 180

```

181 LLSCLTVPASA 191
181 LLSCLTVPASA 191

RESULT 12
POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1 and NS2) (Fragment).
OS Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H., Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes.";
RT Virology 188:331-341(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: D10075; BAA00969.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural
KW INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT CHAIN 347 369
FT TRANSMEM 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448

FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 88.8%; Score 934; DB 1; Length 737;
Best Local Similarity 89.5%; Pred No. 5.1e-73;
Matches 171; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRPPQDVKFGGQIVGVYLLPRGPRGLGVTRATKTSERSNPRG 60
DB 1 MSTNPKPQRLTCRNTNRPPQDVKFGGQIVGVYLLPRGPRGLGVTRATKTSERSNPRG 60
QY 61 RRPNPDKARRDPGRTPWANGYWPPLYGNEGCGWAGLLSPRGSRPSWGPTDWRRSRNLG 120
DB 61 RRPNPDKARRDPGRTPWANGYWPPLYGNEGCGWAGLLSPRGSRPSWGPTDWRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTVPASA 191
DB 181 LLSCLTVPASA 191

RESULT 13
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (SC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (SC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) [RNA-directed RNA polymerase] (EC 2.7.7.48)].
DE Hepatitis C virus (isolate HC-J6) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H., Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions.";
RT J. Gen. Virol. 72:2697-2704(1991).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

FT CHAIN 76 >192 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 192 192
SQ SEQUENCE 192 AA; 20315 MW; 98B48F4C335A84C CRC64;

Query Match 36.6%; Score 385; DB 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 2.4e-26;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 RNLGKVIDTLTCGFADLMGYIPLVCGAPLGGAARALAHGVRLDGVNATGNLPCCSFSI 176
Db 1 RNLGKVIDTLTCGFADLMGYIPLVCGAPLGGAARALAHGVRLDGVNATGNLPCCSFSI 60

QY 177 FLLALLSCLTVPASA 191
Db 61 FLLALLSCLTVPASA 75

RESULT 15
POLG HCVR8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
corresponding to the flavivirus envelope and NS1 proteins and the
pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
lipoprotein envelope. The envelope consists of two proteins:
protein M and glycoprotein E. The nucleocapsid is a complex of
protein C and mRNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X53131; CAA37291.1; --
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD18e062; HCV_NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
Transmembrane; Nonstructural protein.
FT NON TER 1 1
FT CHAIN <1 76 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 321 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;

Query Match 36.6%; Score 385; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 4.1e-26;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 RNLGKVIDTLTCGFADLMGYIPLVCGAPLGGAARALAHGVRLDGVNATGNLPCCSFSI 176
Db 1 RNLGKVIDTLTCGFADLMGYIPLVCGAPLGGAARALAHGVRLDGVNATGNLPCCSFSI 60

QY 177 FLLALLSCLTVPASA 191
Db 61 FLLALLSCLTVPASA 75

Search completed: August 25, 2004, 04:25:40
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 04:09:51 ; Search time 116 Seconds

(without alignments)
519.517 Million cell updates/sec

Title: US-09-713-687A-1

Perfect score: 1052

Sequence: 1 MSTNPKPQLTCRNTNRRPQ.....CSFSIFLLALLSCLTVPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mbc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	998	94.9	191	12	Q68149
2	998	94.9	191	12	Q68113
3	998	94.9	191	12	Q68123
4	998	94.9	191	12	Q68139
5	998	94.9	191	12	Q68124
6	998	94.9	191	12	Q8BES1
7	998	94.9	415	12	Q995G6
8	998	94.9	415	12	Q995G1
9	998	94.9	415	12	Q995G7
10	998	94.9	415	12	Q995G4
11	998	94.9	415	12	Q995G5
12	998	94.9	746	12	Q8JPM2
13	998	94.9	778	12	Q04184
14	998	94.9	3011	12	Q91FE5
15	998	94.9	3011	12	Q36579
16	998	94.9	3011	12	Q36610

17	998	94.9	3011	12	Q9ELS8	Q9els8 hepatitis c
18	998	94.9	3011	12	Q36609	Q36609 hepatitis c
19	998	94.9	3011	12	Q36608	Q36608 hepatitis c
20	994	94.5	191	12	Q68574	Q68574 hepatitis c
21	994	94.5	195	12	Q81SE2	Q81se2 hepatitis c
22	993	94.4	640	12	Q68966	Q68966 hepatitis c
23	993	94.4	3011	12	Q9DIT6	Q9dit6 hepatitis c
24	992	94.3	191	12	Q68111	Q68111 hepatitis c
25	992	94.3	191	12	Q68572	Q68572 hepatitis c
26	992	94.3	415	12	Q995F9	Q995f9 hepatitis c
27	992	94.3	415	12	Q995F7	Q995f7 hepatitis c
28	992	94.3	415	12	Q995F8	Q995f8 hepatitis c
29	992	94.3	415	12	Q995G0	Q995g0 hepatitis c
30	991	94.2	502	12	Q81565	Q81565 hepatitis c
31	989	94.0	191	12	Q68576	Q68576 hepatitis c
32	989	94.0	191	12	Q31SE6	Q31se6 hepatitis c
33	989	94.0	195	12	Q31SE4	Q31se4 hepatitis c
34	989	94.0	195	12	Q91SE5	Q91se5 hepatitis c
35	989	94.0	326	12	Q81257	Q81257 hepatitis c
36	989	94.0	415	12	Q995G3	Q995g3 hepatitis c
37	989	94.0	778	12	Q04185	Q04185 hepatitis c
38	988	93.9	411	12	Q81812	Q81812 hepatitis c
39	987	93.8	191	12	Q8V7Z8	Q8v7z8 hepatitis c
40	987	93.8	191	12	Q8V7Q9	Q8v7q9 hepatitis c
41	987	93.8	191	12	Q8V809	Q8v809 hepatitis c
42	987	93.8	191	12	Q8V7Q3	Q8v7q3 hepatitis c
43	987	93.8	191	12	Q8V7J4	Q8v7j4 hepatitis c
44	987	93.8	191	12	Q8V806	Q8v806 hepatitis c
45	987	93.8	191	12	Q8V7Z3	Q8v7z3 hepatitis c

ALIGNMENTS

RESULT 1

Q68149 ID Q68149 PRELIMINARY; PRT; 191 AA.

AC Q68149;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Genome polypeptide (Fragment).

GN C.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=US11;

RX MEDLINE=94336721; PubMed=8058787;

RA Bukh J., Purcell R.H., Miller R.H.;

RT "Sequence analysis of the core gene of 14 hepatitis C virus

RT genotypes.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=US11;

RX MEDLINE=92279243; PubMed=1317578;

RA Bukh J., Purcell R.H., Miller R.H.;

RT "Sequence analysis of the 5' noncoding region of hepatitis C virus."

RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=US11;

RX MEDLINE=93376778; PubMed=8396266;

RA Bukh J., Purcell R.H., Miller R.H.;

RT "At least 12 genotypes of hepatitis C virus predicted by sequence

RT analysis of the putative E1 gene of isolates collected worldwide."

RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

PROTEIN C AND MRNA (BY SIMILARITY).

EMBL; U10232; AAA21071.1; -.

PIR; PQ0804; PQ0804.

GO; GO:0019028; C:viral capsid; IEA.

GO; GO:0005198; F:structural molecule activity; IEA.

InterPro; IPR002522; HCV_capsid.

InterPro; IPR002521; HCV_core.

Pfam; PF01543; HCV_capsid; 1.

Pfam; PF01542; HCV_core; 1.

Polyprotein.

NON_TER 191 191

SEQUENCE 191 AA; 20766 MW; 4506596F21B4A893 CRC64;

Query Match 94.9%; Score 998; DB 12; Length 191;

Best Local Similarity 96.3%; Pred. No. 1.7e-83;

Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKPQRLTCRNTNRRPQDVKFGGGQIVGGVLLPRRGLGVRATRKTSERSNPRG 60

DB 1 MSTNPKPQRLTCRNTNRRPQDVKFGGGQIVGGVLLPRRGLGVRATRKTSERSNPRG 60

QY 61 RNPPIKARRPDGRTWANGPYWPPLYGNEGCGWAGWLLSPRGRSFGTDPWRRSRNLG 120

DB 61 RNPPIKARRPDGRTWANGPYWPPLYGNEGCGWAGWLLSPRGRSFGTDPWRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYPIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

DB 121 KVIDTLTCGFADLMGYPIPLVGAPLGGAARALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191

DB 181 LLSCLTVPASA 191

RESULT 2

Q68113 PRELIMINARY; PRT; 191 AA.

ID Q68113

AC Q68113

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Genome polyprotein (Fragment).

C.

GN Hepatitis C virus.

OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OC NCBI_TaxID=11103;

OX [1]

RN SEQUENCE FROM N.A.

RP STRAIN=DR4;

RC MEDLINE=94336721; PubMed=8058787;

RA Bukh J., Purcell R.H., Miller R.H.;

RT "Sequence analysis of the core gene of 14 hepatitis C virus

RT genotypes."

RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243 (1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=DR4;

RC MEDLINE=92279243; PubMed=1317578;

RA Bukh J., Purcell R.H., Miller R.H.;

RT "Sequence analysis of the 5' noncoding region of hepatitis C virus."

RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946 (1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=DR4;

RC MEDLINE=93376778; PubMed=8396266;

RA Bukh J., Purcell R.H., Miller R.H.;

RT "At least 12 genotypes of hepatitis C virus predicted by sequence

RT analysis of the putative E1 gene of isolates collected worldwide."

RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238 (1993).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF


```

CC      PROTEIN C AND MRNA (BY SIMILARITY).
DR      EMBL: U10206; AAA21045.1; -.
DR      PIR: PQ0804; PQ0804.
DR      GO: GO:0019028; C: viral capsid; IEA.
DR      GO: GO:0005198; F: structural molecule activity; IEA.
DR      InterPro: IPR002522; HCV capsid.
DR      InterPro: IPR002521; HCV core.
DR      Pfam: PF01543; HCV capsid; 1.
DR      Pfam: PF01542; HCV core; 1.
KW      Polyprotein.
FT      NON_TER 191
SQ      SEQUENCE 191 AA; 20766 MW; 4506596F21B4A893 CRC64;

      Query Match          94.9%; Score 998; DB 12; Length 191;
      Best Local Similarity 96.3%; Pred. No. 1.7e-83;
      Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGRLGVRAIRKTSERSNPRG 60
DB      1 MSTNPKPQRLTCRNTNRRPQDVKFFGGQIVGGVYLLPRRGRLGVRAIRKTSERSNPRG 60

QY      61 RRNPIPKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPMRRSRNLG 120
DB      61 RRQPIPKARRPDGRTWANPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPTDPMRRSRNLG 120

QY      121 KVIDTLTCGFADLMGYIFLVGAPLGGAAALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180
DB      121 KVIDTLTCGFADLMGYIFLVGAPLGGAAALAHGVRLVEDGVNYATGNLPGCSFSIFLLA 180

QY      181 LLSCLTVPASA 191
DB      181 LLSCLTVPASA 191

RESULT 5
Q68124      PRELIMINARY; PRT; 191 AA.
AC      Q68124
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Genome polyprotein (Fragment).
GN      C.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SW1;
RX      MEDLINE=94336721; PubMed=8058787;
RA      Bukh J., Purcell R.H., Miller R.H.;
RT      "Sequence analysis of the core gene of 14 hepatitis C virus
RT      genotypes.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SW1;
RX      MEDLINE=92279243; PubMed=1317578;
RA      Bukh J., Purcell R.H., Miller R.H.;
RT      "Sequence analysis of the 5' noncoding region of hepatitis C virus.";
RL      Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SW1;
RX      MEDLINE=93376778; PubMed=8396266;
RA      Bukh J., Purcell R.H., Miller R.H.;
RT      "At least 12 genotypes of hepatitis C virus predicted by sequence
RT      analysis of the putative E1 gene of isolates collected worldwide.";
RL      Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```



```
QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 8
Q995G1
ID Q995G1 PRELIMINARY; PRT; 415 AA.
AC Q995G1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1/85;
RX MEDLINE=21126427; PubMed=11222726;
RA Lu L., Nakano T., Orito E., Mizokami M., Robertson B.H.;
RT "Evaluation of Accumulation of Hepatitis C Virus Mutations in a
RT Chronically Infected Chimpanzee: Comparison of the Core, E1, HVR1, and
RT NS5b Regions.";
RL J. Virol. 75:3004-3009(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF268575; AAK19133.1; -.
DR PIR; PQ0804;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002531; HCV env.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
DR Polyprotein; Transmembrane.
FT NON_TER 415 415
SQ SEQUENCE 415 AA; 44638 MW; 5BCBF053351165A9 CRC64;

Query Match 94.9%; Score 998; DB 12; Length 415;
Best Local Similarity 96.3%; Pred. No. 4e-83;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKQRLTCRNTNRRPQDVKEFGGQIVGGVYLLPRRGPRLGVRATRTKTSRSNPRG 60
Db 1 MSTNPKQRLTCRNTNRRPQDVKEFGGQIVGGVYLLPRRGPRLGVRATRTKTSRSNPRG 60

QY 61 RRNPDKARRPDGRTWANGPYWPLYGNEGCGWAGWLLSPRGSRPSMGPTDPMRRSRNLG 120
Db 61 RRQPIKARRPEGRWTWQPGYPWPLYGNEGCGWAGWLLSPRGSRPSMGPTDPMRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGLGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGLGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 9
Q995G4
ID Q995G4 PRELIMINARY; PRT; 415 AA.
AC Q995G4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
```

```
Q995G7
ID Q995G7 PRELIMINARY; PRT; 415 AA.
AC Q995G7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1/79;
RX MEDLINE=21126427; PubMed=11222726;
RA Lu L., Nakano T., Orito E., Mizokami M., Robertson B.H.;
RT "Evaluation of Accumulation of Hepatitis C Virus Mutations in a
RT Chronically Infected Chimpanzee: Comparison of the Core, E1, HVR1, and
RT NS5b Regions.";
RL J. Virol. 75:3004-3009(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF268569; AAK19127.1; -.
DR PIR; PQ0804; PQ0804.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
DR Polyprotein; Transmembrane.
FT NON_TER 415 415
SQ SEQUENCE 415 AA; 44694 MW; 611D181A2E5652D7 CRC64;

Query Match 94.9%; Score 998; DB 12; Length 415;
Best Local Similarity 96.3%; Pred. No. 4e-83;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSTNPKQRLTCRNTNRRPQDVKEFGGQIVGGVYLLPRRGPRLGVRATRTKTSRSNPRG 60
Db 1 MSTNPKQRLTCRNTNRRPQDVKEFGGQIVGGVYLLPRRGPRLGVRATRTKTSRSNPRG 60

QY 61 RRNPDKARRPDGRTWANGPYWPLYGNEGCGWAGWLLSPRGSRPSMGPTDPMRRSRNLG 120
Db 61 RRQPIKARRPEGRWTWQPGYPWPLYGNEGCGWAGWLLSPRGSRPSMGPTDPMRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGLGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGLGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTVPASA 191
Db 181 LLSCLTVPASA 191

RESULT 10
Q995G4
ID Q995G4 PRELIMINARY; PRT; 415 AA.
AC Q995G4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
```

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=HCV-1/82;
 MEDLINE=21126427; PubMed=11222726;
 Lu L., Nakano T., Orito E., Mizokami M., Robertson B.H.;
 "Evaluation of Accumulation of Hepatitis C Virus Mutations in a
 Chronically Infected Chimpanzee: Comparison of the Core, E1, HVR1, and
 NS5b Regions";
 J. Virol. 75:3004-3009(2001).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA (BY SIMILARITY).
 EMBL; AF268572; AAK19130.1; -;
 PIR; P00804; P00804.
 GO; GO:0016021; C:integral to membrane; IEA.
 GO; GO:0019028; C:viral capsid; IEA.
 GO; GO:0019031; C:viral envelope; IEA.
 GO; GO:0005198; F:structural molecule activity; IEA.
 InterPro; IPR002522; HCV capsid.
 InterPro; IPR002521; HCV env.
 InterPro; IPR002531; HCV NS1.
 Pfam; PF01543; HCV capsid; 1.
 Pfam; PF01542; HCV core; 1.
 Pfam; PF01539; HCV env; 1.
 Pfam; PF01560; HCV NS1; 1.
 ProDom; PD186062; HCV NS1; 1.
 Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 Polyprotein; Transmembrane.
 NON_TER 415 415
 FT SEQUENCE 415 AA; 44756 MW; 58CTD6884295ADD3 CRC64;
 SQ

Query Match 94.9%; Score 998; DB 12; Length 415;
 Best Local Similarity 96.3%; Pred. No. 4e-83;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 0;

QY 1 MSTNPKPQRLTCNTNRRPQDVKFGGQIVGGVYLLPRGPRGLGVRATKTSERSNPRG 60
 Db 1 MSTNPKPQRTKNTNRRPQDVKFGGQIVGGVYLLPRGPRGLGVRATKTSERSQPRG 60

QY 61 RRPPIKARRPDGRTWANGYPWPLYGNECGWAGWLLSPRGRSPSGPTDPRRSRNLG 120
 Db 61 RRQPIKARRPEGRTWANGYPWPLYGNECGWAGWLLSPRGRSPSGPTDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSLCLTVPASA 191
 Db 181 LLSLCLTVPASA 191

RESULT 11
 Q995G5 PRELIMINARY; PRT; 415 AA.
 AC Q995G5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=HCV-1/81;
 MEDLINE=21126427; PubMed=11222726;
 RX

RA Lu L., Nakano T., Orito E., Mizokami M., Robertson B.H.;
 "Evaluation of Accumulation of Hepatitis C Virus Mutations in a
 Chronically Infected Chimpanzee: Comparison of the Core, E1, HVR1, and
 NS5b Regions";
 J. Virol. 75:3004-3009(2001).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA (BY SIMILARITY).
 EMBL; AF268571; AAK19129.1; -;
 PIR; P00804; P00804.
 GO; GO:0016021; C:integral to membrane; IEA.
 GO; GO:0019028; C:viral capsid; IEA.
 GO; GO:0019031; C:viral envelope; IEA.
 GO; GO:0005198; F:structural molecule activity; IEA.
 InterPro; IPR002522; HCV capsid.
 InterPro; IPR002521; HCV core.
 InterPro; IPR002519; HCV env.
 InterPro; IPR002531; HCV NS1.
 Pfam; PF01543; HCV capsid; 1.
 Pfam; PF01542; HCV core; 1.
 Pfam; PF01539; HCV env; 1.
 Pfam; PF01560; HCV NS1; 1.
 ProDom; PD186062; HCV NS1; 1.
 Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 Polyprotein; Transmembrane.
 NON_TER 415 415
 FT SEQUENCE 415 AA; 44746 MW; B64D181A26F93FA5 CRC64;
 SQ

Query Match 94.9%; Score 998; DB 12; Length 415;
 Best Local Similarity 96.3%; Pred. No. 4e-83;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 0;

QY 1 MSTNPKPQRLTCNTNRRPQDVKFGGQIVGGVYLLPRGPRGLGVRATKTSERSNPRG 60
 Db 1 MSTNPKPQRTKNTNRRPQDVKFGGQIVGGVYLLPRGPRGLGVRATKTSERSQPRG 60

QY 61 RRPPIKARRPDGRTWANGYPWPLYGNECGWAGWLLSPRGRSPSGPTDPRRSRNLG 120
 Db 61 RRQPIKARRPEGRTWANGYPWPLYGNECGWAGWLLSPRGRSPSGPTDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSLCLTVPASA 191
 Db 181 LLSLCLTVPASA 191

RESULT 12
 Q8JPM2 PRELIMINARY; PRT; 746 AA.
 AC Q8JPM2;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=21904749; PubMed=11907246;
 RX Blanchard E., Brand D., Trassard S., Goudeau A., Roingeard P.;
 "Hepatitis C virus-like particle morphogenesis";
 J. Virol. 76:4073-4079(2002).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA (BY SIMILARITY).
 EMBL; AF529293; AAM94419.1; -;
 DR

DR GO: 0016787; F: hydrolase activity; IEA.
 DR GO: 0003723; F: RNA binding; IEA.
 DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: 0008236; F: serine-type peptidase activity; IEA.
 DR GO: 0005198; F: structural molecule activity; IEA.
 DR GO: 0006740; F: transferase activity; IEA.
 DR GO: 0006118; F: electron transport; IEA.
 DR GO: 0006508; F: proteolysis and peptidolysis; IEA.
 DR GO: 0006350; F: transcription; IEA.
 DR GO: 0019079; F: viral genome replication; IEA.
 DR GO: 0019087; F: viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser trypsin.
 DR InterPro: IPR000345; CytC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRP.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR004109; Peptidase C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01538; HCV NS1; 1.
 DR Pfam: PF02907; HCV NS2; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; viral_RdRP; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;
 Query Match 94.9%; Score 998; DB 12; Length 3011;
 Best Local Similarity 96.3%; Pred. No. 3.8e-82;
 Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 Qy 1 MSNPKPQRLTCNTNRPPDVKPPGGQIVGGVYLLPRGPRGLGVATKTSERSNPRG 60
 Db 1 MSNPKPQRLTCNTNRPPDVKPPGGQIVGGVYLLPRGPRGLGVATKTSERSNPRG 60
 Qy 61 RRPPIKARPDGRTWANGVPEWLYNCGGAGWLLSPRGSPSPMGPTDPRWRSNRLG 120
 Db 61 RRPPIKARPDGRTWANGVPEWLYNCGGAGWLLSPRGSPSPMGPTDPRWRSNRLG 120
 Qy 121 KVIDTLTCGFADLMGYIPLVGAIPGGAARALAHGVRLDGVNATGNLPGCSFSIFLLA 180
 Db 121 KVIDTLTCGFADLMGYIPLVGAIPGGAARALAHGVRLDGVNATGNLPGCSFSIFLLA 180
 Qy 181 LLSCLTVPASA 191
 Db 181 LLSCLTVPASA 191
 RESULT 15
 O36579 PRELIMINARY; PRT; 3011 AA.
 AC O36579;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H77;
 RC MEDLINE=97373636; PubMed=9228008;
 RX Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
 Rice C.M.;
 RT "Transmission of hepatitis C by intrahepatic inoculation with
 transcribed RNA.";
 RL Science 277:570-574(1997).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF009606; AAB66324.1; -.
 DR PIR; A44150; A44150.
 DR PIR; P00804; P00804.
 DR PIR; PS0326; PS0326.
 DR PIR; PS0327; PS0327.
 DR PIR; PS0328; PS0328.
 DR HSSP; P27958; 1HEI.
 DR GO: 0016021; C: integral to membrane; IEA.
 DR GO: 0013028; C: viral capsid; IEA.
 DR GO: 0013031; C: viral envelope; IEA.
 DR GO: 0005524; F: ATP binding; IEA.
 DR GO: 0008026; F: ATP dependent helicase activity; IEA.
 DR GO: 0016787; F: hydrolase activity; IEA.
 DR GO: 0003723; F: RNA binding; IEA.
 DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: 0008236; F: serine-type peptidase activity; IEA.
 DR GO: 0005198; F: structural molecule activity; IEA.
 DR GO: 0016740; F: transferase activity; IEA.
 DR GO: 0006508; F: proteolysis and peptidolysis; IEA.
 DR GO: 0006350; F: transcription; IEA.
 DR GO: 0019079; F: viral genome replication; IEA.
 DR GO: 0019087; F: viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRP.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR004109; Peptidase C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01538; HCV NS1; 1.
 DR Pfam: PF02907; HCV NS2; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; viral_RdRP; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327182 MW; E2E0EE809C63C1B9 CRC64;

```
Query Match      94.9%; Score 998; DB 12; Length 3011;
Best Local Similarity 96.3%; Fred. No. 3.8e-82;
Matches 184; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRLTCRNTNRRPQDVKFGGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
   |||||
Db 1 MSTNPKPQRLTCRNTNRRPQDVKFGGGQIVGGVYLLPRRGPRLGVRATRKTSERSNPRG 60
   |||||

Qy 61 RRNPIPKARRPDGRTWANGYFWPLYGNEGCGWAGWLLSPRGSRPSWGPTDFWRRSRNLG 120
   |||||
Db 61 RRQETPKARRPEGRITWAGYFWPLYGNEGCGWAGWLLSPRGSRPSWGPTDFWRRSRNLG 120
   |||||

Qy 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
   |||||
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRLVDGVNYATGNLPGCSFSIFLLA 180
   |||||

Qy 181 LLSCLTVPASA 191
   |||||
Db 181 LLSCLTVPASA 191
   |||||
```

Search completed: August 25, 2004, 04:27:44
Job time : 121 secs

This Page Blank (uspto)